MEMORANDUM
RM-5865-DOT
APRIL 1969

ASSESSING ALTERNATIVE TRANSPORTATION SYSTEMS

James R. Miller III.

This research is supported by the U.S. Department of Transportation, Federal Railroad Administration, Office of High Speed Ground Transportation under Contract 3-0008.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

MEMORANDUM
RM-5865-DOT
APRIL 1969

ASSESSING ALTERNATIVE TRANSPORTATION SYSTEMS

James R. Miller III.

PREPARED FOR:
U. S. DEPARTMENT OF TRANSPORTATION
FEDERAL RAILROAD ADMINISTRATION
OFFICE OF HIGH SPEED GROUND TRANSPORTATION

The RAND Corporation
SANTA MONICA · CALIFORNIA

Published by The RAND Corporation

PREFACE

This Memorandum is one in a series being prepared by The RAND Corporation for the Northeast Corridor Transportation Project, Office of High Speed Ground Transportation, U.S. Department of Transportation. The overall research effort is directed toward development of comprehensive and systematic methodology for evaluating the potential utility of alternative transportation proposals.

The series of Memoranda can be classified into several types of papers. In the first and major part we are attempting to integrate our efforts into an overall evaluation framework.* Part II is composed of supporting Memoranda on some of the more relevant theoretical aspects associated with combining many dimensions in an alternative selection process. Part III consists of important gap-filling papers and background materials.

This Memorandum is one of three thus far dealing with techniques for analysis of multi-dimensional alternatives.** The paper concentrates on developing and justifying a systematic procedure for assessing the worth of alternative transportation systems. An example applying this procedure to assessing passenger attributes can be found in RM-5869-DOT.* It is intended that the assessment procedure developed herein gradually will become an important component in a comprehensive regional transportation systems analysis capability. This would be useful in evaluating alternative mixes of transportation modes both at present and in the future.

---

*Frederick S. Pardee, et al., Measurement and Evaluation of Transportation System Effectiveness, The RAND Corporation, RM-5869-DOT, 1969 (forthcoming).

**The other two Memoranda are RM-5877-DOT, Improving the System Design and Evaluation Process by Use of Trade-off Information: An Application to Northeast Corridor Transportation Planning by K. R. MacCrimmon, The RAND Corporation, April 1969; and RM-5868-DOT/RC, Preferences for Multi-Attributed Alternatives by Howard Raiffa, The RAND Corporation, April 1969.

SUMMARY

This paper addresses itself to the problem of assessing worth. It is assumed that a decision context has been specified and that a fixed set of discrete alternatives has been produced. Thus, the decision context might be provision of more rapid passenger service between New York and Washington, and produced alternatives might include a high-speed train, improvements to the highway network connecting the two cities, and an underground tube.

Having specified the decision context and produced several alternatives, it then remains to predict whatever performance would be delivered by each alternative, if implemented; to estimate the total resources of various types required to implement each alternative; to assess the worth of each alternative as viewed by diverse interest groups; to trade off these considerations, along with considerations of risk/uncertainty; and, finally, to arrive at a terminal decision. The bulk of this paper is directed toward worth assessment, although the other problems mentioned above are discussed briefly.

To aid in the assessment and decision process, a systematic procedure has been devised. The purpose and scope of the procedure are set forth in Sections I through V of this paper. Section VI presents the procedure in outline form. Section VII discusses in some detail how the procedure might actually be implemented. Section VIII contains a complete example of the procedure at work, although in a simpler, non-transportation context. Sections IX and X extend the procedure from worth assessment to final decisionmaking. The paper closes with a critical review in Section XI.

An experiment was performed to test the procedure (i.e., to demonstrate that it could be carried out successfully by professional, governmental decisionmakers). Results drawn from the experiment are interpreted, and conclusions are reported in Appendix U.

The procedure is quantitative throughout. However, it relies heavily upon subjective inputs from responsible decisionmaking personnel. The major thesis of the procedure is that assessment and final choice

must depend upon subjective evaluations, but that a systematic and quantitative method of making such judgments proves quite helpful. Reported experimental results support this point of view.

Briefly, the procedure is as follows. Assuming that a decision context (i.e., a job to be accomplished) has been specified and that discrete alternatives have been produced, the first step is to define explicitly what is <u>desired</u> in the way of performance from each alternative. This does <u>not</u> mean predicting what each alternative can <u>deliver</u>. Predicting performance is an entirely separate task. Rather, it means listing overall objectives or major performance criteria and insuring that the list is:

1. Complete (i.e., contains all criteria which decisionmakers are able and willing to formulate and display);
2. Mutually exclusive (i.e., contains criteria which neither encompass nor are encompassed by other criteria on the list);
3. Of major significance (i.e., contains only highest-level criteria); and
4. Free of worth interdependence (i.e., contains only worth-independent criteria).

The meaning of "major significance" and "worth-interdependence" will be clarified in Sections VI and VII.

Having established a list of overall performance objectives, the second step is to generate a hierarchical structure of successively more specific performance criteria. This involves breaking down or subdividing higher-level criteria into one or more lower-level criteria alleged to be included within the meaning thereof. The process of subdivision continues until the decisionmaker becomes confident that he has specified in detail all of the objectives he really possesses. If the decisionmaker thinks of additional objectives during the assessment process, these are appended to the hierarchical criterion structure. If he discovers additional ramifications of existing criteria, then the structure is further subdivided. Both of these events occur frequently during the assessment process. In fact, the procedure has been designed specifically to induce such events and to guide the consequent revisions of the criterion structure.

Once a satisfactory criterion structure has been achieved, the third step is to select a physical performance measure for each lowest-level criterion. Performance measures describe what an alternative can deliver, while performance criteria state what the decisionmaker desires. The purpose of selecting performance measures is to establish concrete connections between desires (existing in the minds of decisionmakers) and deliverable performance from real alternatives. Each performance measure serves to interpret its corresponding lowest-level criterion in physical terms. Thus, if a decisionmaker desires rapid passenger service (a performance criterion), this might be interpreted as average daily trip time in minutes between two specified locations (a physical performance measure).

However, merely establishing interpretive connections is not sufficient in itself to permit formal evaluation. Specific worth relationships must be mapped out between each lowest-level criterion and its related performance measure. This constitutes the fourth step. It is implemented by defining scoring functions which assign a unique numerical worth score to every possible value of a performance measure. Assigned worth scores provide a quantitative indication of whether or not and the extent to which a real alternative satisfies (through its deliverable performance) the decisionmaker's desires with respect to the related criterion. Scoring functions will be defined, either explicitly or implicitly, to provide such quantitative indications of worth for every lowest-level criterion and its related performance measure.

The fifth step is to combine worth scores assigned on the basis of separate performance criteria to arrive at an overall index of each alternative's worth. This is accomplished by defining a weighting function. An additive function with constant trade-off weights will be adopted for this purpose. The reason for choosing an additive function is entirely pragmatic. Previous attempts by real decisionmakers to formulate trade-off relationships among criteria in more complicated ways just plain failed to generate comprehensible results. However, feasibility is purchased at a definite cost. Use of an additive weighting function requires that sets of sub-criteria located at

every branch of the hierarchical criterion structure contain members relatively independent of one another in the worth sense. Again, the meaning of "worth independence" will be clarified in Sections VI and VII.

The sixth step is to validate both the scoring functions and the weighting functions against whatever alternatives have been produced. This means computing an overall worth index for each alternative and judging the results for reasonableness. After having progressed through the lengthy process of assigning scores and weights, decisionmakers typically develop quite strong notions of what constitute reasonable results. In fact, the major virtue of the assessment procedure lies in its ability to clarify within the minds of decisionmakers just what they want from an alternative and just how they might discriminate operationally among alternatives which do and which do not deliver desirable performance.

Results generated during early passes through the procedure are usually unreasonable in some respect, particularly where complex alternatives (such as transportation systems) are involved. Thus, the next step is to make selective revisions in any one or more of the following ways.

1. Additional criteria, suggested in attempting to account for unreasonable results, may be added to the hierarchy. This is by far the most common form of revision.
2. Existing criteria may be further subdivided, re-defined, or completely eliminated. This is also a common form of revision.
3. Scoring functions may be re-calibrated.
4. Weights may be adjusted either to reflect revised notions of the relative importance of satisfying various criteria or to reflect the differential interpretive quality of various performance measures.

A reasonable assessment structure can usually be achieved after several (e.g., four or five) passes. The last step is to trade off each alternative's overall worth with the following considerations in arriving at a final decision:

1. Risk/uncertainty;
2. Required resource expenditures;
3. Temporal changes in objectives, aspiration levels, and tastes; and
4. Different and possibly conflicting points of view among diverse interest groups.

The reader may find the following suggestions helpful in reading this paper. If only a general understanding of the purpose and scope of the procedure and an outline of the assessment process is desired, omit Sections VII and XI and all of the appendices. If detailed knowledge of how to implement the procedure is also desired, read Sections VII and XI and Appendices A through T. If the reader is also interested in the experimental test of the procedure, read Appendix U.

ACKNOWLEDGMENT

The author would like to express his appreciation to F. S. Pardee for recognizing the potential usefulness of my research on problems of multi-dimensional choice analysis and suggesting its extension and application in the field of transportation system evaluation. I am also indebted to several other members of the RAND staff for their suggestions during many working sessions and to S. H. Dole and F. Roberts for their specific comments on a draft version of the memorandum.

CONTENTS

I. INTRODUCTION

Assessing the worth of "complex" alternatives in an "important" decision situation is generally regarded as difficult. However, this task constitutes but one phase in the still more difficult process of producing such alternatives and making a final decision among them. Some of the factors which make an alternative "complex," which make a decision "important," and which thereby render the overall decision process difficult are stated and illustrated in the paragraphs that follow.

Consider the situation faced by the Department of Transportation (DOT) in selecting an alternative transportation system for the Northeast Corridor. Producing, evaluating, and finally selecting an alternative to implement this decision is decidedly difficult for several reasons.

First, it must be determined which activities are to be included within the scope of a transportation alternative. This requires the decision maker to describe in some detail the job to be performed by whichever alternative is finally selected. In the case of a transportation system, such a job description would ordinarily be both lengthy and complex.

Second, having formulated an adequate job description, the next task is to stipulate the overall purpose in selecting a transportation alternative. But transportation systems are multi-purpose rather than single-purpose entities. They satisfy many different objectives simultaneously. For this reason, their overall worth from DOT's point of view cannot be reckoned on the basis of a single criterion. This suggests the need for responsible decision-making personnel to undertake the following tasks.

1. The several objectives which are to be satisfied by acquiring and utilizing a transportation system should be listed, and the list should be fairly complete.
2. From each listed objective should be derived a set of specific worth criteria in terms of which the physical performance of a system may be assessed.

3. Some means should then be found to organize and integrate these multiple criteria into a consistent and meaningful assessment structure.

Third, both the acquisition and the operation of transportation equipment have many important ramifications for various local economies and the national economy. There is no simple or unique consequence on the basis of which an entire decision can be made. There are many performance consequences which must first be ascertained with reasonable accuracy and then assessed meaningfully before a final decision can be reached. This suggests the need for a clear, systematic, and replicable procedure to insure that no important performance consequences are overlooked.

Fourth, since both multiple worth criteria and multiple performance consequences are present, some means should be found to establish connections between the two. But this is not as easy as it may seem. A single performance consequence may be related simultaneously to several worth criteria (e.g., passenger travel time might be considered relevant simultaneously to whether or not the passenger can keep a business appointment at his destination, how long he must suffer some form of discomfort along the way, and when he can retire to bed at the end of his business day). Conversely, several performance consequences may be related simultaneously to a single worth criterion (e.g., air pollution, noise levels, and income redistribution might all be considered relevant measures of societal impact).

Fifth, complex patterns of interaction may exist among various performance consequences due to the fact that transportation systems are both large and highly interrelated. This makes it difficult to understand and, therefore, to predict accurately an entire set of specific consequences. This is particularly troublesome when alternatives are in the design phase of their development.

Sixth, even if all performance consequences were known for certain, there may still exist complex patterns of interaction among the various worth criteria imposed by human beings upon this known performance. The structure of human worth notions is itself infested with

intricate patterns of interdependence. Even worse, human beings often find it exceedingly difficult to distinguish in their own minds between interaction among performance consequences (a physical characteristic of transportation systems) and interdependence among imposed worth criteria (a psychological characteristic of human beings). This renders still more difficult any attempt to understand, to assess, and to select transportation systems.

Finally, there is the question of resources expended to acquire and operate a system. It is not always easy to predict with accuracy the amounts of manpower, materiel, and monetary resources which must be expended in order to implement any proposed alternative. Even if the resource implications of each proposed alternative were predictable, it must still be decided how much of each type of resource <u>should</u> be expended on the particular job under consideration. In other words, the relative importance of the job should be ascertained in advance, and this should be translated into specific amounts of each type of resource which might appropriately be expended to perform that job.

Historically, decision makers have attempted to cope with the above kinds of problems largely on the basis of subjective judgment and intuition. Subjective estimates have been used quite frequently to predict probable resource and performance consequences. Personal judgments have also been used both to assess the worth of different amounts of predicted performance and to effect trade-offs among various worth criteria. The twin problems of physical interaction among performance consequences and conceptual interdependence among worth criteria have been handled similarly--that is, on an intuitive basis. Now if the decision problem under consideration were simple, and if the consequences of making a poor decision were relatively inconsequential, this might be the best way to proceed. The extra gains realizable from formalizing and systematizing the decision process would probably not justify the extra time, cost, and effort required. However, when the problem becomes complex (e.g., in the senses just described), and/or when the consequences become important, strict reliance on unstructured subjective judgment becomes a dangerous gamble indeed. It seems

unreasonable to permit such decisions to be made in the absence of factual evidence, logical discipline, and at least an opportunity to attain consensual validation. This suggests the need for a systematic procedure.

The central purpose of this paper is to develop an explicit, logically consistent, and replicable procedure to aid in evaluating alternative systems.* In addition, the results of an experiment designed to measure the impact of the procedure upon professional decision makers will be reported.** Since almost all of these professional decision makers were U.S. Government employees, their reactions should be of particular interest to DOT officials.

---

*It should be made clear that explicitness, logical consistency, and replicability do <u>not</u> preclude the use of subjective judgment. Quite to the contrary, it is the writer's view that subjective judgment <u>must</u> be used both in assigning measures of worth to various performance consequences and in trading off worth among various criteria. Subsequent sections of this paper will be specifically devoted to supporting this point of view. Rather, what is being stipulated here is that, when used, subjective judgment should be made explicit, should be thoroughly scrutinized for logical consistency, and should be elicited by a uniformly applicable and replicable procedure. The writer can think of no better way to insure that personal judgments will be free of false assumptions than by stating these assumptions explicitly. Nor can he think of a better way to insure valid reasoning from assumptions to conclusions than by exposing the reasoning process to critical scrutiny. Nor can he think of a better way to elicit a cross-section of opinion and to establish a consensus of preferences than by means of a uniformly applicable and replicable procedure. Most important of all, the writer can think of no better ways than these to obtain feedback on the assessment process and, therefore, to provide a constant impetus to its improvement.

**This is not to be confused with the experimental application of the procedure in assessing passenger attributes which is described in detail in Section D-X of RAND RM-5869-DOT. As stated, the experiment referenced in the text above was designed primarily to test various psychological and operational impacts of the procedure on its participants and was conducted prior to the writer's involvement with issues in transportation system evaluation.

II. STATEMENT OF THE PROBLEM

The overall problem, of which this paper will treat only a part, is seven-fold. Assuming that an important decision is to be made among complex transportation alternatives, then the problem is:

1. to <u>describe</u> the job to be performed by whichever alternative is finally selected (i.e., to list the various activities which are to be carried out);

2. to <u>formulate</u> the overall purpose in making the decision (i.e., to abstract a specific set of job objectives from the job description);

3. to <u>produce</u> one or more feasible alternatives (i.e., to design and/or to solicit proposals for at least one alternative whose performance, if selected, would be viewed as satisfying minimum adequacy requirements);

4. to <u>predict</u> the worthwhile performance consequences associated with each alternative (i.e., to predict the types and amounts of worthwhile performance which would be realized from acquiring and utilizing each of the alternatives produced);

5. to <u>assess</u> the worth of these predicted performance consequences (i.e., to assess the extent to which the above-predicted performance would succeed in accomplishing stated job objectives);

6. to <u>predict</u> the resource consequences associated with each alternative (i.e., to predict the types and amounts of limited resources which would necessarily be expended to acquire and utilize each of the alternatives produced);

7. to <u>reach</u> a final decision (i.e., to match worth of performance received against limited resources expended on each of the alternatives produced so as

to determine whether any of them should be selected
and, if so, which one).

Actually, this paper addresses itself almost exclusively to the problem of worth assessment (see 5. above). It will henceforward be assumed for discussion purposes that a job has been described, that an overall decision purpose has been formulated, that one or more feasible alternatives have been produced, and that the physical performance of each alternative has been adequately predicted. Nevertheless, despite these simplifications, the residue of the problem is still very difficult. To illustrate the remaining difficulties, let us consider a hypothetical example.

Suppose that the job is to convey passengers and freight between New York and Philadelphia. Suppose, further, that the only performance consequences considered important are the average line-haul travel time for passengers and the average daily throughput in freight. Suppose, also, that the only resource considered important is the initial dollar investment required to procure the transportation system (i.e., operating costs are ignored). Finally, suppose that three alternative systems have been proposed and that validated estimates of their performance and cost consequences are as shown in Table 1.

Table 1

Performance and Cost Consequences	Alternative I	Alternative II	Alternative III
Trip time	20 minutes	40 minutes	75 minutes
Daily freight capacity	10,000 tons	100,000 tons	75,000 tons
Investment cost	$110 million	$125 million	$90 million

The decision maker would now be faced with the task of making trade-offs between different levels of trip time and freight capacity to arrive at some notion of the overall worth of each alternative, and then he would have to match overall worth against cost on all three.

Comparing alternatives I and II, he would have to decide whether the increase in trip time from 20 to 40 minutes were at least offset by the increase in maximum freight capacity from 10,000 to 100,000 tons. If no, then it certainly would not be worthwhile spending the extra \$15 million to purchase alternative II. If yes, if the increased freight capacity more than compensated for the longer trip time, then he would have to decide whether the net gain in worth derivable from selecting alternative II over I warranted spending the additional \$15 million. But what about alternative III? It is much cheaper than the other two, its freight capacity falls between the other two, and its trip time is substantially longer (and, therefore, less desirable) than both of its competitors'. Comparisons similar to those described above would have to be made first between alternatives I and III, and then between alternatives II and III.

However, the above types of comparisons, even if carried out successfully, would not be sufficient to dispose of the problem completely. There still remain the twin dangers of "under-kill" and "over-kill." One or more of the three alternatives would provide "under-kill" if there existed either a maximum acceptable trip time or a minimum required freight capacity, and if estimated performance fell beyond either of these limits. This is an obvious kind of danger which can usually be detected with little difficulty--particularly if such mandatory performance requirements have been stipulated in advance on the basis of careful engineering and design considerations.

In contrast, the other kind of danger--the danger of "over-kill" --is far more subtle and much more difficult to detect. The reason is that "over-kill" is an economic rather than an engineering concept. Assessment of "over-kill" requires simultaneous consideration of both performance and resource consequences. The essence of "over-kill" lies not in the mere fact that more performance may be proposed than is necessary, but rather in the fact that whatever additional performance (over and above minimum requirements) is proposed may not warrant expending whatever additional resources are required to receive that additional performance. On economic grounds, it may be preferable to accept lesser performance--or even to accept zero performance (i.e.,

abandon the project)--and to expend the saved resources on some other project entirely. Returning to our example, it may be that alternative III, even with its relatively long (75 minutes) trip time, is more than adequate to meet the job requirements. Under such circumstances, it might be economically unwise to spend any more than $90 million on this job. Alternatively, it may happen that even $90 million is too much to spend. It may be that a small improvement to the current system costing only $25 million would also be adequate, and that even the cheapest of the proposed new alternatives (costing $90 million) would not justify spending the extra $65 million. That same money might better be spent on some other transportation project or, perhaps, on some other project completely unrelated to transportation. Before a final decision can be made, all of these issues should be considered, and the decision maker should be prepared to reject any (or even all) of the proposed alternatives if either "under-kill" or "over-kill" becomes apparent.

The difficulty of making the above types of trade-off decisions--first between different kinds of performance to arrive at an assessment of overall worth, and then between overall worths and their associated costs--is probably quite evident to the reader. And this was a highly simplified example. As the number of worth criteria and related performance consequences increases, the problem quickly reaches unmanageable proportions. If multiple resources are also considered (e.g., manpower and materiel as well as monetary resources), and if complex patterns of both physical interaction and worth interdependence emerge, then effective solution of the problem by unstructured intuition becomes just about impossible. This suggests the need for a more formal approach.

The remainder of this paper will be oriented specifically toward the development of a more formal approach. Sections III through VIII will set forth a systematic procedure to aid in the assessment of worth. Section IX will extend the results of the procedure to produce monetary as well as non-monetary measures of worth. Section X will present a technique for comparing alternatives dynamically (i.e., over time) as well as statically (i.e., a fixed point in time). A technique

for representing different and frequently conflicting points of view will also be presented. Section XI will close with a critical summary. In addition, Appendix U reports the results of an experiment designed to test the procedure as implemented by professional decision makers.

III. THE CONCEPT OF WORTH

For purposes of this paper, the worth of any object, activity, or situation is, roughly speaking, the extent to which such is perceived by a decision maker or group of decision makers as satisfying clearly articulated objectives. Thus, the worth of an alternative in a stated job context would be defined in terms of how well that alternative satisfied whatever objectives have been established regarding the job to be accomplished.

The above notion of worth is intentionally stated in very general terms. A detailed definition will be presented later. Specifically, step-by-step procedures for assessing worth will be outlined in Sec. VI and developed more fully in Sec. VII. One purpose in setting forth these procedures is to provide an operational definition of the concept itself. For now, however, it will be useful to outline the intended meaning and scope of the worth concept--both to orient future discussion and to preclude the imputation of unintended meanings to the subject matter of this paper.

THE INTENDED MEANING OF WORTH

From the above discussion, it is apparent that worth notions constitute an internal property of human decision makers--not an external property of the physical objects, activities, and situations whose worth is being assessed. Worth is here conceived as inherent within the perceptual apparatus of the decision maker himself. The detailed procedures to be developed in Secs. VI and VII will clarify this distinction operationally.

Since worth is defined with respect to clearly stated objectives, it is necessary that such objectives exist. Operationally, this requires that a deliberate effort be made to formulate and articulate clear objectives before worth may be assessed. It also means that worth notions will be multidimensional whenever multiple objectives and/or multiple performance measures are considered relevant (e.g., in the case of "complex" alternatives).

In addition, worth refers to the extent or degree to which some object, activity, or situation satisfies stated objectives. This suggests the need for establishing a definite scale in terms of which various degrees of goal satisfaction (and, therefore, imputed worth) may be expressed. Section IV will address itself to establishing such a scale.

IMPLICATIONS FOR THE TASK OF ASSESSING WORTH

Having discussed briefly the meaning of worth, as it will be used in this paper, we shall now investigate some of its implications for the practical problem of assessing alternatives.

First, the act of formulating and articulating a clear set of objectives in terms of which worth may be assessed is not always easy to accomplish. Decision makers may be either unable or unwilling to formulate and display a complete list of objectives because of:

1. incomplete awareness of the problem at hand;
2. incomplete knowledge of the intricacies of the problem; or
3. inability (due to time, money, and/or manpower constraints) to devote sufficient "thinking" effort to formulating a complete and explicit list of objectives.

Alternatively, they may be unwilling to formulate and particularly to display a complete list of objectives because of:

1. fear that some of the "real" objectives will be disapproved if laid bare to public scrutiny;
2. fear that some of the "real" objectives, even if tacitly approved, may not be easily defended in the political arena; or
3. realization that some objectives, even if approved and defensible, may not receive complete consensual validation from all interested parties--particularly those who would suffer adverse consequences should the "real" objectives be satisfied.

These latter sources of unwillingness may attain particular motivational importance if decision makers are themselves imbedded in an organizational environment rife with threat, conflict, or a strong tradition of defensive conservatism.

Second, there is the issue of confirming worth judgments. Unlike allegations of fact or scientific predictions, worth judgments cannot be confirmed by empirical test. They are <u>in principle</u> untestable by ordinary scientific means. This is because worth judgments are stated in such a way as to be neither factually true nor factually false. They merely exist in the minds of human beings to be accepted or rejected either in whole or in part by other human beings (or, perhaps, by the same human being at a different point in time). In short, the acceptability of worth judgments is here conceived as a matter of informed opinion.

A third implication follows from the second. This involves the identity of decision makers. Different decision makers may very well have different objectives regarding the same situation, which renders the outcome of an assessment highly dependent upon who undertakes to perform that assessment. Stated a bit more simply, the outcome of an assessment depends critically upon <u>whose</u> values are adopted. One way out of this situation is to strive for consensus among potential decision makers, but this is not always possible (and perhaps even undesirable). In any case, the worth concept is not here <u>defined</u> as requiring consensus.

A fourth implication involves the stability of worth judgments over time. Not only may there exist lack of consensus among separate decision makers at a given point in time, but there may also exist lack of agreement among separate worth judgments made by the same decision maker at different points in time. As additional experience is gained, one would expect (or at least hope) that a given decision maker would alter his worth judgments to account for whatever new insights this additional experience has brought about. Temporal instability is thereby created, but, possibly, in an entirely appropriate manner. In any case, the worth concept is not here <u>defined</u> as requiring temporal stability either.

IV. CONSTRUCTING A MEASURE OF WORTH

Having committed ourselves to creating a formal assessment scheme, we must now tackle the problem of defining a uniform and convenient measure of worth. This suggests (although it does not require) reducing the problem to numbers. Why? Because numbers are familiar, widely used as tools of measurement, and easy to manipulate. However, lest there be any confusion on this issue, let it be understood at the outset that the measure of worth to be created, and particularly its numerical scale characteristics, constitute an ad hoc invention specifically designed for our assessment procedure. No claim is being made that this worth measure or its scale characteristics derive deductively from any set of logical or mathematical axioms.

THE BASIC PURPOSE IN CONSTRUCTING A MEASURE OF WORTH

Perhaps the best way to initiate detailed discussion is with a formal statement of purpose. When a numerical measure of worth is used as a vehicle of assessment, the underlying rationale is that worth numbers or worth scores will be assigned such that numerical relationships among assigned worth numbers will faithfully reflect perceived worth relationships among the objects and activities to which these numbers have been assigned. In order to implement this purpose, it is first necessary to specify the kinds of worth relationships which are to be reflected by means of numerical symbols. It is also necessary to specify the numerical conventions which establish a correspondence between numerical relationships and the perceived worth relationships which are being depicted thereby. The first of these tasks will be undertaken in the next section. The second will be undertaken in the section, Corresponding Scale Characteristics.

WORTH CHARACTERISTICS TO BE REFLECTED

The most fundamental characteristics of the worth concept to be reflected in our choice of a numerical measure are the three psychological states of preference, aversion, and genuine indifference.

A decision maker is said to possess a postive preference for some object or activity if and only if that object or activity elicits a positive affective response from him (e.g., joy, pleasure, interest, excitement, gratification, etc.). Thus, most people possess a positive preference for automobiles because they elicit all of the above positive responses. In addition, most people are willing to part with money (a scarce resource) in order to receive these benefits from an automobile.

A decision maker is said to possess a distinct aversion (negative preference) for some object or activity if and only if that object or activity elicits a negative affective response from him (e.g., distress, anxiety, shame, guilt, disgust, etc.). Most people possess a distinct aversion to death. The very thought of it arouses a great deal of distress and anxiety, and many people are willing to part with substantial amounts of money (i.e., purchase life insurance) in order to ameliorate its unwanted consequences.

A decision maker is said to feel genuinely indifferent toward some object or activity if and only if he possesses neither a preference for nor an aversion toward that object or activity.

Returning to the concept of worth, this is usually thought of as related to positive preferences only. That is, when an object or activity is said to possess some worth, this usually means that somebody possesses a positive preference for it and/or its consequences. The concept of "negative" worth (referring to objects or activities toward which people feel aversive) is less well defined.

In light of these observations, certain numbers on the worth scale (to be created in the next section) will be reserved to indicate positive preferences, and a single number not included in the above range will be reserved to indicate a state of genuine indifference. Negative preferences or aversions will be represented by negative analogs of the positive worth numbers.*

*More will be said in subsequent sections about negative preferences or aversions and their representation on the worth scale.

Another aspect of the worth concept to be reflected in our choice of a numerical measure involves its boundedness. Is it possible for something to be completely worthwhile? Can a decision maker be completely satisfied (or dissatisfied) with some object or activity? Although seemingly simple at first glance, this is a very subtle and important question. Let us investigate the issue more closely.

If asked to assess the worth of some object without specifying how or for what purpose that object is to be used, it seems difficult to conceive of any natural, logical outer bounds to the answer. Like the brightness of a color or the loudness of a sound, there exist no apparent natural limits. On the other hand, once a definite job has been specified and once a definite set of objectives has been defined, then the question appears in a somewhat different light. When asked to assess the worth of some object for performing some stated job in accordance with well defined objectives, it seems reasonable to talk in terms of the extent to which that object satisfies the stated objectives. Furthermore, since definite objectives have been stated, it seems reasonable to talk about the possibility, at least, of having those objectives completely satisfied. Thus, under these revised circumstances, there appears to be a natural outer bound to the assessment of worth. This will be reflected by placing numerical bounds on the worth scale.

Still another aspect of the worth concept to be represented numerically is its continuity or divisibility. It would seem desirable to permit the expression of preferences and preference differences to range from infinitesimal magnitudes to large magnitudes. Although decision makers may not always wish to avail themselves of this flexibility, a continuous or everywhere-dense worth scale will be defined to accommodate such notions whenever they are felt.

Finally, there is the question of preference relationships between different objects or activities whose worth is being assessed. There are three kinds of basic relationships which will receive numerical representation by establishing appropriate scale conventions. These are:

1. same-difference relationship (i.e., whether two objects or activities are assessed as possessing the same or different worths);
2. greater than and less than relationships (i.e., whether one object or activity is assessed as possessing more or less worth than another);
3. comparative magnitude relationship (i.e., how many times as much worth one object or activity is assessed as possessing compared to another).

Let us now proceed to construct a numerical worth scale which will reflect all of the above characteristics.

CORRESPONDING SCALE CHARACTERISTICS

A numerical worth scale will be established in accordance with the following ten scaling conventions.

1. Positive numbers will be assigned uniformly to situations assessed as possessing positive worth (i.e., toward which a positive preference is felt).
2. Negative numbers will be assigned uniformly to situations assessed as possessing "negative" worth (i.e., toward which a distinct aversion is felt).
3. The worth scale will be bounded from above by plus one and from below by minus one.
4. Plus one will be assigned only to those situations deemed completely successful in terms of accomplishing positive job objectives. Analogously, minus one will be assigned only to those situations deemed completely "successful" in accomplishing "negative" job objectives (i.e., to situations than which nothing worse is conceivable in the context of the stated job).
5. The number zero will be assigned uniformly to situations assessed as completely worthless (i.e., completely unsatisfactory--but not dissatisfactory; toward which genuine indifference--but not aversion--is felt.
6. All real numbers between plus one and minus one (inclusive) are permissible measures of worth.
7. Two situations will be assigned equal worth numbers if and only if they are assessed as possessing identical worth (i.e., a decision maker feels genuine indifference in choosing between them).

8. One situation will be assigned a higher worth number than another if and only if it is assessed as possessing more worth--that is, if and only if a decision maker prefers the first situation to the second.

9. Numbers between zero and plus one (exclusive) will be assigned to all situations assessed as partially successful in terms of accomplishing positive objectives. Worth numbers will be assigned to such situations according to their proportional or percentage accomplishment of the stated objectives. This defines magnitude comparisons in terms of their ratios.

10. Numbers between zero and minus one (exclusive) will be assigned to all situations assessed as partially "successful" in terms of accomplishing "negative" objectives (i.e., stated avoidance desires). Negative worth numbers will be assigned to such situations according to their proportional or percentage "accomplishment" of stated "negative" objectives.

CHOICE OF A UNIT OF WORTH: WORTH POINTS (GRATILES)

The choice of a unit of worth has already been made implicitly by two previous decisions. First, it was decided to bound the worth scale from above and below (i.e., to restrict worth numbers to fall between plus and minus one). This precludes the use of dollars or any other unit whose range lacks intrinsic logical outer bounds.

Second, it was decided to assign a worth number to a situation according to the proportional accomplishment of stated objectives achieved by that situation. This means that worth numbers may be viewed as ratios between actually achieved satisfaction and maximum possible satisfaction of stated objectives. As such, no matter what units raw satisfaction might possess, any ratio formed would be dimensionless. That is, such raw units would cancel each other out in forming a ratio, and the result would be dimensionless. Worth numbers defined in this manner are like index numbers used by various rating schemes (e.g., The Consumer Price Index).

Despite their lack of physical units, there is still an advantage to giving these index numbers a definite label so that they may be easily remembered and conveniently discussed. Henceforward, our worth numbers will be referred to as "points," "worth points," "gratification

points," or "gratiles," where all four labels are understood to be completely synonymous and interchangeable. No matter which label is used, however, their significance is encapsulated within the ten scaling conventions set forth in the above section.

LEGITIMATE OPERATIONS ON WORTH POINTS

Before attempting to develop a procedure for assessing worth, it would be wise to conclude this section with a statement of legitimate operations that may be performed on worth points. The legitimacy of these operations follows from the ten scaling conventions previously outlined.

First, assuming that the process of assigning worth points is adequate to convey the meaning encapsulated in the ten scaling conventions, then all four basic arithmetic operations may be performed upon them, except for a sign restriction. Worth points may be added, subtracted, and multiplied by a non-negative constant with complete freedom. However, attention must be paid to their sign when multiplying or dividing by one another. Only worth numbers of like sign may undergo these operations, and then only their absolute magnitude is relevant. In the language of scaling theory, each half of the worth scale constitutes a full-fledged ratio scale, with the negative half being treated as a "mirror-image" of the positive half.

A second issue concerns the legitimacy of assigning worth points to situations toward which no positive preference is felt. Ignoring the case of indifference, which receives a point score of zero, this includes both situations toward which a distinct aversion is felt and situations toward which neither a positive preference nor a distinct aversion are felt directly, but whose indirect consequences are such as to arouse a reduction in positive feelings.

An example of a situation generating direct aversive feelings would be the development of a very high-speed train which was both cheap and comfortable, but which occasionally exploded upon encountering an obstruction on the tracks. Otherwise positive attitudes toward the rapid, comfortable, and cheap ride would be substantially mitigated, if not completely overruled by the fear of explosion, maiming, and

possible death. Avoidance of these undesirable side effects would constitute a "negative" objective toward which most people would feel a distinct aversion. Negative worth points would be assigned to these side effects according to their likelihood of occurrence and to the degree of ensuing disaster, should they occur.

However, there is another type of situation toward which decision makers may feel neither a direct preference nor a direct aversion. This is the situation wherein limited resources must be expended to complete a job. Unless a decision maker possesses miserly feelings, he has no direct aversion to spending money or committing workers or using up capital equipment per se. If the supply of such resources were truly unlimited relative to their demand, the resources themselves would have no worth at all--either positive or negative. Consequently, spending such unlimited resources could only be regarded with indifference. There would always be enough to go around--if the supply were truly unlimited. Resources only become valuable when their available supply falls below the total demand for their effective utilization. But even then, their worth is not intrinsic to the resources themselves. Rather, their worth derives from the fact that they may be diverted to some alternative application which, if carried out, would generate consequences perceived as worthwhile in their own right. Expending limited resources to complete one job precludes using the same resources to complete some alternative jobs, and the worth of completing the alternative jobs must, therefore, be forgone.

In view of these observations, we may now ask whether it is legitimate to assign worth points directly to the expenditure of resources. The answer is: not usually. Worth points, as defined in this paper, can only be assigned to situations perceived as worthwhile in their own right because they succeed in accomplishing stated objectives. Although it would be possible to define "conserving resources" as a specific objective, it would be difficult to judge the worth of any given amount of conserved resources unless or until the alternative applications of the same resources had first been ascertained and assessed. Until this is accomplished, no meaningful point scores may be assigned to resource expenditures.

The above conclusions have two important procedural implications. Since worth points are generally not assigned to resource expenditures incurred in acquiring and utilizing a produced alternative, while worth points are assigned to other kinds of consequences related directly to stated job objectives, it is important to define at the outset just which consequences are to be regarded as resource-oriented and to distinguish these clearly from objective-oriented consequences. In addition, some alternative means must be found to reflect the worth implications of expending resources and to incorporate these explicitly into an overall decision making methodology. Section V will discuss briefly various ways of incorporating resource considerations into an overall decision making methodology. Section X will extend this discussion substantially, ending up with a procedure to convert worth point scores into equivalent resource units (e.g., dollars).

V. RELATED CONCEPTS

Before moving to the development of a formal assessment procedure, the relationship of worth to the classical concept of utility deserves some brief attention. In addition, the roles of worth, utility, and resources in the overall decision making process deserve a few brief comments. Although neither of these topics will be treated extensively within this section, a brief discussion of each will add perspective to our future discussions of worth assessment.

RISK, UNCERTAINTY, AND THE CLASSICAL CONCEPT OF UTILITY

The worth concept is completely devoid of any risk and/or uncertainty considerations. In assessing the worth of a situation, activity, or performance consequence, it is assumed that such an outcome will occur <u>for certain</u>. Consequently, assigned worth numbers will <u>not</u> reflect the aversion which a decision maker may feel toward either risk or uncertainty regarding the actual occurrence of that outcome. Furthermore, the process of assigning worth numbers provides no mechanism for reflecting perceived trade-offs between the worth of an outcome, conditional upon its actual occurrence, and the variable risk or uncertainty surrounding its occurrence. The worth concept and the related worth-measuring and worth-assessing procedures are, therefore, incomplete in this sense.

In contrast, the classical concept of utility, as articulated by Von Neumann and Morgenstern and used by statistical decision theorists, does provide an explicit mechanism for reflecting perceived trade-offs, but it ignores the problem of formulating and articulating a measure of conditional worth. It assumes that the decision maker has already formulated a worth measure and proceeds from there.

A complete assessment procedure should take account of both conditional worth and risk/uncertainty considerations. That is, it should provide a mechanism for assessing worth, conditional upon certainty, and then it should provide an additional mechanism to account for the decision maker's attitudes toward risk and uncertainty. While this

paper will address itself exclusively to the former task, powerful extensions (and some revisions) of our procedure have been developed by Howard Raiffa.* Raiffa's procedures have the distinct advantage of introducing explicit uncertainty considerations into every step of the assessment process. Hence, Raiffa's measures of worth may be combined with performance probabilities to arrive at expected utilities of various alternative systems. Whether Raiffa's procedures prove easier or more difficult to implement "in the field" must be determined by practical decision makers.

THE CONCEPT OF A DECISION RULE

A decision rule might be defined broadly as any uniformly applicable directive which indicates a clear choice among properly specified alternatives in a given decision situation. The principal role of a decision rule is to provide an explicit vehicle through which decision makers may express their willingness to make trade-offs among worth, risk/uncertainty, and resource considerations.**

Two examples of frequently used decision rules are:

1. The Economy Rule, directing decision makers to select the least expensive feasible alternative (i.e., the least resource-consuming alternative which satisfies all stipulated mandatory performance requirements and, possibly, physical resource limitations); and

2. The Ratio Optimizing Rule, directing decision makers to select whichever feasible alternative maximizes a utility-to-cost (or, equivalently, minimizes a cost-to utility) ratio.***

Obviously, the above list does not exhaust all decision rules that have

---

*See Howard Raiffa, Preferences for Multi-Attributed Alternatives, The RAND Corporation, RM-5868-DOT/RC, April 1969.

**The trade-off between conditional worth and risk/uncertainty is assumed throughout this discussion to be encapsulated in a utility index of the variety discussed above.

***The word "cost" is used throughout this section to indicate the physical process of expending resources including, but not restricted to, monetary resources.

been or could be used to select alternatives, but it does provide a reasonable basis for discussion. In particular, it provides a reasonable basis for illustrating the primary role of a decision rule in integrating worth, risk/uncertainty, and resource considerations.

In choosing a decision rule, the decision maker must ask himself what he is really trying to accomplish when he finally selects an alternative. He may raise such questions as the following:

1. Assuming that at least one of the produced alternatives is feasible, must one of them always win the selection; or is it possible to reject <u>all</u> of the alternatives on the grounds that they <u>all</u> provide "over-kill" and that the same resources might better be expended on some other project altogether?

2. Should each successive selection in which the decision maker is required to make a choice be considered separately, without regard to the consequences of that choice on subsequent selection decisions, or should the decision maker assume a broader viewpoint which embraces the <u>whole sequence</u> of decisions he must make?

3. In what sense should valuable performance received be compared with resources expended? Is it worthwhile to expend additional resources in order to receive additional valuable performance over and above minimum requirements? If so, how much more and until what has been achieved?

Answers to these questions should help the decision maker choose a decision rule, or at least narrow substantially the field of candidates. To illustrate why this is so (i.e., how these questions are related to various decision rules), let us consider the implicit answers given to each by the economy rule and the ratio optimizing rule.

First, the economy rule requires that, if at least one feasible alternative is produced, then one of them must win. It is impossible, under the economy rule, to reject all feasible alternatives--even if the least costly alternative requires a staggering expenditure of resources. No protection against "over-kill" is provided.

Similarly, the ratio optimizing rule (in either of its two equivalent forms) provides no protection against "over-kill." It is quite possible to encounter a set of alternatives--all of which promise performance greatly in excess of what is required (or even desired) and

which involve commensurately excessive resource expenditures. Nevertheless, that alternative with optimum ratio would still be defined and, unless the rule were enhanced with a budget constraint or some other protective device, "over-kill" would thereby be suffered.

Regarding the second question, both the economy rule and the ratio optimizing rule focus attention exclusively on each successive selection considered by itself. No explicit consideration is given to the consequences of one selection decision on other such decisions. In particular, no recognition is given to the fact that, when the total supply of resources is limited (as it almost always is in real-world situations), what must be expended to choose an alternative in one selection cannot be expended on another selection. No limits or any other direct controls are placed on resource expenditures.

Regarding the third question, the two decision rules give quite different answers. The economy rule rejects completely the notion that additional performance over and above minimum requirements might be worth spending additional resources to obtain. It chooses the cheapest alternative that does the job, even if performance is just barely satisfactory.

In contrast, the ratio-optimizing rule recognizes the potential worth of additional performance over and above minimum requirements, and it permits spending additional resources to obtain it. Under this rule, the goal is to obtain the "best buy" (i.e., the most for whatever resources are expended as evidenced by a maximum or minimum ratio).

The preceding discussion was intended to indicate that, even after a satisfactory measure of worth has been defined, and even after risk/uncertainty has been taken into account by means of a satisfactory utility index, there still remains the problem of integrating these considerations with a careful consideration of resource expenditures before a complete decision methodology can be achieved. Choice of a satisfactory decision rule constitutes a means of achieving integration, and, as the preceding discussion illustrated, this is not a simple task with an obvious solution. We shall return to the question of selecting an overall decision rule in the latter sections of this paper.

VI. OUTLINE OF A PROCEDURE FOR FORMULATING AN ASSESSMENT STRUCTURE

In Section I of this paper it was pointed out that worth assessment is an especially difficult task in the case of "complex" alternatives due to:

1. Multiple objectives and assessment criteria to list and arrange in some organized fashion;

2. Multiple performance consequences to predict;

3. Multiple worth connections between listed assessment criteria and predicted performance consequences;

4. Physical interaction among performance consequences; and

5. Worth interdependence among assessment criteria.

However, the difficulty of this task can and will be reduced somewhat by making two simplifying assumptions. First, it will be assumed that validated estimates are freely available for all relevant performance consequences associated with all produced alternatives. Naturally, both obtaining and validating such estimates constitute very real and highly important problems in their own right, but neither of these will be discussed here in any detail. Such omissions are purely for simplification.

Second, it will be assumed that our task is restricted to assessing a fixed set of discrete alternatives. The problem of producing alternatives (i.e., of designing, redesigning, or soliciting proposals for alternatives) will not be considered. This assumption reduces substantially any worries we might otherwise have had concerning physical interaction among performance consequences, since physical interaction is troublesome primarily because it renders prediction of performance difficult under differing design alternatives.*

*The scope of our problem is greatly reduced by this assumption, but not to the point where it no longer possesses practical significance. After all, in any real decision situation, there comes a moment when the process of producing alternatives must be terminated, and an immediate choice must be made among whichever alternatives have already been produced. At that moment of decision it is reasonable to view the choice as among a fixed set of discrete alternatives.

Nevertheless, in spite of these simplifications, we must still worry about listing and arranging multiple objectives and assessment criteria, checking for worth interdependence among them, and establishing worth connections between these criteria and various performance consequences. The remainder of this section will address itself to these tasks.

LISTING OVERALL PERFORMANCE OBJECTIVES

The first step in making a formal assessment is to specify what is desired from whatever alternatives may be produced. This means listing objectives. At the outset, objectives may be (and should be) stated in very general terms. After all, the point is to be as all-encompassing as possible initially (to avoid omitting any important objectives which decision makers really possess and are willing to display), and then to work down through a process of successive elaboration to a very specific statement of desired performance. A very specific statement of intentions is required at the end of the process in order to carry out an actual assessment, but this need not concern us too heavily at the beginning. Rather, what should concern us is summarized below.

Any list of overall performance objectives should possess the following desirable properties.

1. The list should be complete and exhaustive. That is, all important performance objectives deemed relevant to the final decision should be represented by the items on the list. This is to guarantee that no important performance considerations are overlooked by the assessment procedure.

2. The list should contain mutually exclusive items. That is, no listed objective should be stated in such a way as to encompass (definitionally) or to be encompassed by (definitionally) any other objective either in whole or in part. This is to permit decision makers to view listed objectives as

independent entities among which appropriate trade-offs may be established. This will also help prevent undesirable "double-counting" in the worth sense.

3. The list should be restricted to performance objectives of the highest degree of importance. That is, only overall objectives should be included. The purpose of this exclusion is to provide a sound basis or starting point from which lower-level criteria may subsequently be derived.

4. Finally, the list should contain objectives relatively independent in the worth sense. That is, for any pair of objectives on the list, decision makers should be willing to trade-off additional satisfaction on one objective for reduced satisfaction on the other at a rate relatively independent of the level of satisfaction already attained on each. The meaning of independence and the reasons for requiring objectives to be independent will be discussed in Identifying and Eliminating Worth Dependence Among Separate Performance Criteria.

GENERATING A HIERARCHICAL STRUCTURE OF PERFORMANCE CRITERIA

Having established a list of overall performance objectives satisfying the above four logical requirements, the second step is to define more precisely what these highest-level objectives really mean. To accomplish this, each highest-level objective is subdivided into one or more lower-level criteria. The purpose of subdividing is to state explicitly (i.e., in terms of lower-level criteria) what is intended by or included within the meaning of each overall objective. But what, exactly, is the nature of this task?

Essentially, our task is to create a pictorial map of the structure of worth relationships residing within the mind of a decision maker. Just as a cartographer attempts to depict topographical relationships of distance, elevation, contiguity, etc., between masses of land and

water in some specified geographical region, we are trying to depict worth relationships among overall performance objectives and successively lower levels of increasingly more specific performance criteria relevant to the selection of a specified alternative for some definite job. Just as the cartographer utilizes certain conventions such as countour lines and special coloring to convey information about the terrain he is describing, we shall adopt the convention of a tree-like array to convey information about the decision maker's worth structure.

Despite these similarities, however, there are a number of important differences between constructing maps of regional topography and maps of human worth structure. First, the cartographer attempts to describe various aspects of our physical environment. We, on the other hand, are attempting to describe various aspects of the inner minds of human decision makers. This suggests that the proper focus of our attention is <u>not</u> the "out-there" physical world of nature, but rather the "in-here" subjective world of human beings. It is to decision makers and their evaluative responses that we must look in constructing our map.

A second difference follows immediately from the first. Since the cartographer is attempting to map something physical and directly observable, he may utilize direct measuring devices such as compasses and other surveying tools. We, on the other hand, are attempting to map something non-physical and only indirectly measurable. We are therefore forced to utilize indirect measuring devices such as verbal questioning and behavioral observation. From these kinds of data we must <u>infer</u> the underlying structure of human preferences.

A third difference relates to the number and temporal stability of the entities being mapped. Whereas there is only one topographical region to be investigated by the cartographer (the particular region he is interested in mapping), there is frequently more than one decision maker to be investigated in mapping a worth structure (e.g., the group of decision makers responsible for making a selection decision). In addition, topographical features of our physical environment are apt to be highly stable over time, while attitudinal features of our assessment structure are apt to change over time with new learning and increased assessment experience.

A fourth difference, and by far the most important one, relates to the perturbing effect of the mapping process itself. The cartographer is concerned with depicting visually a territorial region which has already been formed by the forces of nature. His mapping process does not alter significantly the nature of the physical terrain being mapped. In sharp contrast, our mapping process has an enormous impact upon the worth structure being mapped. On the basis of an experiment (reported in Appendix U) it was concluded that the single most important consequence of the entire assessment procedure is to create a worth structure where one did not previously exist--at least not in conscious, well-defined, and easily articulatable form. Participation in this assessment procedure induces the decision maker to formulate a consistent worth structure. At the very least, this entails substantial clarification of what already existed in his mind. Typically, it induces him to alter substantial portions of his prior worth structure. At most, it induces him to create a structure which did not enjoy any prior existence at all in consciousness. Producing a pictorial map of the decision maker's worth structure, once formulated, constitutes a separate and important consequence of the assessment procedure, but this is not the only consequence, nor is it the most important one.

More will be said later about the dynamic interrelationship between formulating and representing a worth structure. For now, however, we shall concentrate primarily upon the representational or mapping aspects of the process. By means of a step-by-step questioning procedure, a hierarchical, tree-like structure of increasingly more specific performance criteria is generated to represent what the decision maker desires from produced alternatives.

SELECTING PHYSICAL PERFORMANCE MEASURES

After generating a hierarchical tree of overall objectives and lower-level criteria, the third step is to select a single physical performance measure for each lowest-level criterion on the tree. The purpose of selecting physical performance measures is to give concrete, physical interpretations to their related lowest-level criteria. By

this device, a bridge is constructed linking the subjective inner mind (i.e., the worth structures) of decision makers to the objective outer world of physical alternatives. Let us clarify this concept--particularly the distinction between performance criteria and physical performance measures--with further discussion.

A physical performance measure is any tangible reading or concrete observation that can be extracted from the real world. For purposes of assessment, it is any directly measurable attribute of a produced alternative. However, this is not the same thing as a performance criterion. Whereas stated criteria reflect what a decision maker desires from produced alternatives, performance measures reflect what an alternative can actually deliver. Performance criteria are attributes of decision makers, while performance measures are attributes of the physical alternatives being assessed.

Although this may sound like a mere academic distinction, it will be useful for very practical reasons to keep the two concepts clearly separated. There are two reasons for maintaining the distinction. First, the methods of approach and the people one talks to in formulating performance criteria are different from the methods and people involved in defining physical performance measures. Introspective reflection and discussions with fellow decision makers can help to formulate, to clarify, and to understand performance criteria. This seems like a reasonable way to define what is desired from an alternative. In contrast, inspection of physical alternatives and discussion with knowledgeable engineers would seem a more useful way to define physical performance measures. These reflect what an alternative will deliver (no matter what is desired).

A second reason for distinguishing between performance criteria and performance measures springs from the very different way in which they will be treated in the process of formal assessment. Once defined, physical performance measures will be used to describe each of the produced alternatives. The description of an alternative in terms of a set of physical performance numbers (and/or other descriptive symbols) will then be converted into equivalent worth point scores by means of a device called a scoring function (to be discussed in the next section).

In direct contrast, worth scores attached by scoring functions to lowest-level performance criteria are not themselves run through scoring functions. Instead, they will be combined with other worth scores already attached to other performance criteria. Such combination will be carried out by means of a device called a weighting function (to be discussed later), and the result will be a single, overall index of worth associated with each produced alternative.

Selecting physical performance measures must be done judgmentally. A decision maker must choose a well-defined and easily measurable physical attribute of an alternative which he feels serves to interpret, in phenomenological terms, the intended meaning of the lowest-level criterion under consideration. Thus, returning to the transportation example, the performance criterion "daily freight throughput" might be interpreted by the physical measure, "maximum daily tonnage which could be transported over the line-haul link joining New York and Philadelphia, assuming no accidents, breakdowns, or other mishaps." But this raises two questions. First, how does one come up with a candidate measure? Second, if more than one candidate arises, how does one choose among them?

Coming up with candidate measures, just like generating subcriteria to fill out the hierarchical tree, requires ingenuity and informed judgment. Both tasks involve creative acts. However, both tasks will be much easier to accomplish if decision makers take the trouble to compile in advance a master list of reasonable candidates. This master list might contain all performance criteria and all performance measures that have ever been suggested and/or used on past decisions of a similar nature. Particular criteria and measures could then be extracted (or synthesized) from the master list as needed for each successive decision. In addition, the master list could be continually updated to include new criteria and measures as they are created.

As for choosing among candidate measures, this also requires informed judgment. It may happen, for example, that certain modes of transportation (airplanes) are known to suffer frequent delays (when airport traffic patterns become jammed), while other modes (trains)

rarely suffer the same kind of delays. Under such circumstances, a better measure of "daily freight throughput" might be "expected daily tonnage" with delays and other mishaps taken into consideration by means of historical frequency data describing the delays typically suffered by each mode. This kind of choice among candidate measures must be made by decision makers on the basis of historical evidence and their experience in making assessments.

ESTABLISHING SPECIFIC WORTH RELATIONSHIPS BETWEEN LOWEST-LEVEL PERFORMANCE CRITERIA AND THEIR ASSOCIATED PHYSICAL PERFORMANCE MEASURES: THE SCORING PROBLEM

The fourth step in constructing a formal assessment procedure is to establish specific worth relationships between each lowest-level performance criterion in the hierarchical structure and its associated physical performance measure. Selecting performance measures (the step just discussed in the above section) serves to establish the existence of worth connections, but it does not serve to map out specific worth relationships. Specific relationships are established by means of scoring functions.

A scoring function is a mathematical rule which assigns a unique worth score in points to every possible value of some physical performance measure. It transforms raw performance (measured in terms of whatever physical unit is appropriate to the performance measure under consideration) into worth-of-performance (measured in terms of the worth points discussed in Section IV). Just as the selection of a physical performance measure serves to interpret concretely each lowest-level performance criterion and, therefore, to provide a bridge between the subjectively defined worth structure and the objectively defined physical characteristics of an alternative, the specification of a scoring function serves to define precisely the nature, shape, and particular parameters of this bridge.

In formulating a scoring function, it is temporarily assumed that the lowest-level performance criterion in question constitutes the only performance objective in the entire assessment. Then, the worth score assigned by the scoring function to any given amount of performance

on the associated physical performance measure is supposed to indicate the extent to which that amount of physical performance actually satisfies the lowest-level criterion. To accomplish this, certain conventions or ground rules must be observed uniformly to insure that all worth scores thereby generated will be comparable with one another and subject to a uniform interpretation. Otherwise, the subsequent procedure by which individual worth scores assigned to separate criteria are to be combined cannot be meaningfully carried out. A set of scoring conventions designed to insure both consistency and comparability appears below.

1. The outputs of all scoring functions will be in terms of worth points.

2. Worth points will be as defined in accordance with the ten scaling conventions presented in Section IV.

3. All scoring functions will be formulated to cover the <u>entire</u> range of logically possible physical performance--not just the reasonable or expected range. This is to insure that a definite point score will be defined for every conceivable level of produced performance--not matter how unexpected it may be.

4. Most scoring functions will take the form of mathematical formulas and/or graphically depicted mathematical curves. However, some will not be expressed in these terms. Some will take the form of direct judgmental point assignments by decision makers without the aid of either formulas or graphs. In this latter case, scoring functions are thought of as implicit in the minds of the decision makers.

5. All scoring functions will be formulated by means of a single, uniform, and replicable procedure. A suggested two-stage procedure (embodying the above four scoring conventions) will be presented in Sec. VII and illustrated in Sec. VIII.

COMBINING WORTH SCORES ASSIGNED ON THE BASIS OF SEPARATE PERFORMANCE CRITERIA TO ARRIVE AT A SINGLE, OVERALL INDEX OF WORTH: THE WEIGHTING PROBLEM

In discussing scoring functions, it was suggested that one temporarily assume each lowest-level performance criterion under consideration to be the sole performance objective in the entire assessment. Obviously, this is an untenable assumption. There are many performance objectives to be satisfied as reflected in the hierarchical structure and its many lowest-level branches. This brings us to the fifth step in formal assessment--combining worth scores assigned on the basis of separate performance criteria to arrive at a single, overall index of worth. This step will be accomplished by defining a weighting function.

A weighting function is a conceptual device by means of which explicit recognition is given to the existence of multiple objectives and performance criteria. Whereas a scoring function is defined to indicate the extent to which any given level of measured performance succeeds in satisfying its related lowest-level performance criterion, a weighting function is defined to indicate the perceived relative importance of satisfying the criterion itself compared with other performance criteria. In this manner, the temporary assumption of a single criterion made in defining a scoring function is relaxed to reflect reality. Simultaneously, a means of combining worth scores assigned on the basis of separate criteria into a single, overall index is achieved. Let us illustrate these results by means of a very simple example.

Suppose that a new transportation system is to be acquired and operated between Boston and Washington with two specific objectives in mind. These are:

1. to carry the currently existing load of passengers between Boston and Washington; and
2. to expand at some future date so as to accommodate part of the passenger load between intermediate points (e.g., New York, Philadelphia, and Baltimore).

Suppose, also that performance of the current job is to be measured by maximum daily passenger volume between Boston and Washington and

that an appropriate scoring function has been defined to convert daily volumes into equivalent worth scores. Finally, assume that expansion potential is to be measured by maximum additional passenger volumes between intermediate cities and that a scoring function has also been defined for this performance measure. Then, each alternative system would be assigned two worth scores--one for performing the current job and another for demonstrated expansion potential. How can these two separate scores now be combined into an overall index of each alternative's total worth? This is the weighting problem.

One way to proceed would be as follows. Decision makers ask themselves which of the two performance criteria--doing the current job or providing expansion capability--should be considered more important. That is, if given the choice between satisfying either of the two criteria to the same extent, which one would they prefer to have satisfied? If decision makers would prefer to have the current job criterion satisfied over having the expansion potential criterion satisfied to the same extent, then the former criterion must be considered more important than the latter. If genuine indifference is felt between having the two criteria equally well satisfied, then they must be regarded as equally important.

The next step is to be a bit more precise about the extent or degree of perceived relative importance. Just to say that doing the current job is more important than providing expansion potential is usually not sufficient to distinguish clearly between the overall worths of competing alternatives. The magnitude of this perceived relative importance must also be indicated. How much more important is it to satisfy the current job criterion than to satisfy the expansion potential criterion? Twice as important? Ten times as important? Representation of relative magnitudes once again suggests resorting to numbers.

Suppose that performing the current job were considered three times as important as providing expansion potential. Then, any pair of numbers standing in the ratio of 3:1 could be used to convey this information. In particular, the numbers 3/4 and 1/4 could be used.

Then, whatever scores are attached by scoring functions to these two criteria could be combined by:

1. multiplying the score assigned to performing the current job by 3/4;

2. multiplying the score assigned to expansion potential by 1/4; and

3. adding the two products to arrive at a weighted average score, using the importance ratios as constant weights.

The resulting sum of weighted scores might then be interpreted as an overall index of each alternative's total worth.

The above procedure has a definite appeal in its simplicity and directness. It seems to solve the problem of combining scores on separate criteria, and it seems to arrive at a single, overall index of worth. Moreover, by requiring the set of constant weights to add internally to one (as was done in the example above), the resulting overall worth score (computed as the sum of weighted individual criterion scores) also lies between zero and one and may be subjected to the same interpretation as worth point scores assigned to individual performance criteria. This renders far more manageable the task of checking assigned weights for intuitive reasonableness and consensual validation. The same questions may be asked of weighted sums as are asked of individual criterion scores. Since worth scores cannot be validated by any other means (recall that they are in principle untestable by ordinary scientific techniques), uniform interpretability becomes an extremely important and valuable asset.

However, in spite of its simplicity and immediate appeal, the above procedure should be subjected to critical scrutiny before accepting it and incorporating it into a formal assessment scheme. It would be wise to inquire a bit more carefully into what this weighting procedure is really assuming about how decision makers view multiple assessment criteria, how they trade off worth among multiple criteria, and what procedural implications these assumptions have for the practical task of assessment. It will be shown that the key to understanding these issues lies in the concept of worth interdependence among separate

performance criteria. This concept and its procedural implications will be discussed in the next section.

IDENTIFYING AND ELIMINATING WORTH INTERDEPENDENCE AMONG SEPARATE PERFORMANCE CRITERIA

The preceding example of combining worth scores by means of weighting and summing to arrive at a single index of overall worth assumes implicitly the following things:

1. The relative importance of satisfying separate performance criteria does not depend upon the various degrees to which each criterion has itself been satisfied. Rather, their relative importance is conceived as being constant in this respect.
2. The rate at which increased satisfaction of any given criterion contributes to overall worth is independent of the levels of satisfaction already achieved on that and other criteria. Rather, such rates are viewed as constant in this respect.
3. The rate at which decision makers would be willing to trade off decreased satisfaction on one criterion for increased satisfaction on other criteria so as to preserve the same overall worth is independent of the levels of satisfaction already achieved by any and all of the criteria. Such trade-off rates are viewed as constant in this respect.

These three logically interrelated statements, taken together, define the concept of worth independence.

To clarify further this concept of worth independence and, more particularly, to distinguish it from interdependence, let us consider two contrasting examples. First, we shall return to the example given in the preceding section and argue that performing the current job and providing expansion potential constitute worth-independent criteria. Then, we shall concoct a counter-example to illustrate worth-interdependence.

The two criteria previously considered were:

1. ability to handle the current passenger volume between Boston and Washington; and

2. expandability of the system at some future date to transport passengers between intermediate points.

These two criteria are independent because decision makers would like to have both satisfied simultaneously and because the extent to which either may be satisfied does not depend upon the extent to which the other has already been satisfied. Under these circumstances, it seems reasonable to combine criterion scores by means of constant relative importance weights.

Now let us concoct an example of substantial worth interdependence. Suppose that every proposed alternative includes both a propulsion mechanism and a passenger-carrying compartment. Looking at the problem through the eyes of a system design engineer, it might seem reasonable to define the following two criteria:

1. performance of the propulsion mechanism; and

2. performance of the passenger-carrying compartment.

Now these would not be independent criteria from the passenger's point of view. They would be highly interdependent. Why? Because passengers would be unimpressed with the most beautifully designed propulsion mechanism if they were forced to endure a bumpy, hot, smelly, and uncomfortable ride. Similarly, passengers would be unimpressed with the most luxurious and comfortable accommodations if the vehicle continually broke down en route. Passenger satisfaction with the performance of either depends critically upon satisfaction with the performance of the other. Stated alternatively (and more incisively), passengers care little about the independent performance of either; their real concern lies with the joint performance of both components acting together as a unit.

The foregoing example serves not only to illustrate the concept of worth independence, it also provides an excellent basis for outlining some of the common ways in which interdependencies arise and for

underscoring the subtle but important distinction between perceived worth interdependence and actual performance interaction. Let us pursue these topics in some detail.

The reader may have concluded that worth interdependence arose above from the engineering difficulties in designing a propulsion mechanism without considering the design characteristics of the passenger compartment, or vice versa. Although such difficulties would certainly arise, this is not the point at all. This illustrates performance interaction--a severe problem facing any system design engineer--but it does not illustrate worth interdependence--an assessment problem facing passengers and decision makers charged with taking the passenger's point of view. The essential distinction lies in whose problem and whose point of view is being taken. While performance interaction is the design engineer's problem, and solving such problems constitutes one of his important objectives, this is not the passenger's problem. The passenger is concerned only with enjoying solutions to the engineer's problem (i.e., riding on whatever vehicle is eventually designed and produced). Interdependencies among factors contributing to overall passenger satisfaction do not necessarily correspond to interdependencies among factors contributing to the design and production of the vehicle. To keep these concepts separate, we refer to the first as worth interdependencies and to the second as performance interactions.

One further point. If we consider the design engineer's problem, then worth interdependence and performance interaction achieve conceptual correspondence. This is because his objective is to design and to worry about the production of a total vehicle, but not to ride in it. Hence, worth interdependence among design criteria does arise from corresponding performance interactions from his point of view. This brings us back to a statement made several times earlier in this paper. Any assessment procedure, to generate comprehensible results, must stipulate very clearly whose point of view is being taken and whose values are to prevail. This will not guarantee consensus, but it should improve clarity.

One disturbing consequence flows from the above distinction. Detailed knowledge of the design and engineering factors incorporated in a transportation system constitutes an inadequate and frequently misleading basis for assessing its worth to passengers, other beneficiary groups, and society at large. The assessment task requires decision makers to empathize with the values, attitudes, and perceptions of ultimate users. Engineering and related forms of technical training might actually interfere with this process. By the same token, knowledge and training in financing, administering, and regulating transportation systems could interfere with the assessment process unless utilized with great care.

Returning to the concept of worth interdependence, and realizing that it exists only in the minds of ultimate beneficiaries and decision makers charged with assessment, we can now discuss some of the common ways in which it arises. Several of these are listed below:

1. part-whole interdependence, where ultimate beneficiaries are not concerned with distinct parts of a whole system or subsystem, but only in the attributes of the whole entity they form when properly combined (this form of interdependence was just illustrated);

2. means-end interdependence, where ultimate beneficiaries do not attach worth to alternate means of achieving the same final end, but only to the manner and degree to which the final end is achieved (thus, providing train and plane service between two cities would normally not constitute two independent objectives); and

3. dominating factor interdependence, where some factor, if present, serves to dominate ultimate beneficiaries' perception of an alternative (thus, high death probabilities would render virtually irrelevant all other considerations in selecting an alternative system).

Our assessment procedure, to be developed more fully in Sec. VII, contains several devices to identify and purge worth interdependence from the criterion hierarchy. Restricting the hierarchy to contain

only independent criteria will permit free use of constant trade-off weights. Past research has shown that decision makers find it difficult to conceive of trade-offs except in this relatively simple way.

THE MEANING AND INTERPRETATION OF WEIGHTS

Just as it was useful to establish by means of explicit scale conventions the meaning and proper interpretation of worth point scores, so also is it useful to establish a similar logical basis for numerical weights. This will be accomplished by stating and discussing briefly ten weighting conventions.

1. A set of numerical weights will be defined for every set of sub-criteria into which a higher-level criterion in the hierarchical criterion structure is subdivided. In the case of the highest-level of overall performance objectives, these are construed as "sub-criteria" of "overall worth" and, therefore, each of these will also receive a numerical weight. In all cases, a single weight will be defined for each such sub-criterion.
2. The numerical weight attached to each sub-criterion will be interpreted as an indication of the perceived relative importance of satisfying that sub-criterion in the context of the higher-level criterion within whose meaning it is alleged to be included. Relative importance means "relative to the other sub-criteria in the set."
3. Relative importance will be reflected in the <u>ratios</u> of any two weights assigned, respectively, to two separate sub-criteria in a given set. It is in such ratios that trade-off rates will be embodied.
4. Weights will be assigned <u>only</u> to sub-criteria perceived as <u>devoid</u> of substantial worth interdependence. A definite procedure has been devised to identify and eliminate sub-criteria displaying substantial worth interdependence.

5. Weights will be restricted to fall within the range of non-negative numbers. This is to indicate that the concept of relative importance possesses only "positive" connotations. Restricting weights to fall within the range of non-negative numbers guarantees that all trade-off rates (i.e., all ratios between pairs of weights) will be non-negative.
6. Theoretically, a weight of zero would be assigned to any sub-criterion in a given set of sub-criteria if and only if satisfying that sub-criterion were perceived as completely unimportant. In practice, however, a sub-criterion to which a zero weight might appropriately be attached will be ignored (i.e., such a sub-criterion will <u>not</u> be included in the hierarchical criterion structure), since, by the above definition, its satisfaction is viewed as totally unimportant. This definition is included only to provide a logical lower bound to the range of permissible weight numbers and to give the lower bound a definite interpretation.
7. All of the weights in any given weight set (corresponding to a given set of sub-criteria) will add to a finite positive constant, and the same positive constant will apply to all weight sets. This serves to normalize assigned weights so that a given weight number will always have the same significance (i.e., indicate the same relative importance) in all weight sets. Consequently, the task of validating weight assignments by visual inspection becomes easier.
8. The finite positive constant to which all weights in any given weight set add will be one. Any such constant would be permissible, but setting this number equal to one has a certain conceptual appeal. Since all weights are non-negative and add to one, each weight must lie between zero and one. Hence, relative importance may be viewed as if it were a percentage or proportion, which

decision makers may find to be a convenient and familiar conceptual aid.

9. Assigned weight numbers cannot exceed one, and a weight of exactly one will only be assigned in cases where a set of sub-criteria contains a single member. Then, that single sub-criterion must receive full weight. As such, it must be interpreted as equivalent in the worth sense to its related higher-level criterion.
10. Any positive real number equal to or less than one will be a permissible weight. This will permit the formation of any desirable trade-off ratio by properly selecting pairs of weights.

ADJUSTING THE WEIGHTS TO REFLECT THE RELATIVE INTERPRETABILITY OF EACH PHYSICAL PERFORMANCE MEASURE

Another issue, which has not yet been discussed, concerns the relative extent to which each physical performance measure previously selected to interpret (in physical terms) its associated lowest-level performance criterion does in fact succeed in providing an adequate interpretation thereof.[*] Decision makers might view "expected daily tonnage hauled" as a good way to measure the lowest-level criterion "freight throughput." This is because it reflects very well the intended meaning of "freight throughput" in the context of the transportation job under consideration. In contrast, "total number of discrete promises" found in a formal proposal submitted by a locomotive manufacturer to perform that job might be considered a poor measure of "manufacturer's good faith." This is because "manufacturer's good faith" refers to an attitude on the part of corporate executives, and this attitude may not be clearly reflected in the text of their formal proposal. Discussions with executives and review of their

---

[*]The writer is indebted to H. Martin Weingartner for originally raising and noting the importance of this issue. The writer is also indebted to Howard Raiffa for criticizing constructively the particular manner in which this issue is treated in the assessment procedure.

historical behavior in similar contractual situations should provide vastly superior measures of their good faith.

To the extent that wide differences emerge in the relative interpretive quality of various performance measures, this could have a seriously distorting impact upon the outcome of a decision. It is quite conceivable that a relatively important criterion (deserving a large numerical weight) cannot be interpreted with any measures of good quality because the decision maker is unable to articulate in explicit physical terms what he means by this criterion. The decision, therefore, should not be unduly influenced by such criteria, especially if other criteria--even though considered relatively less important--are much easier to interpret in terms of high-quality measures. In short, there should be some explicit mechanism for reflecting the relative quality of each criterion's interpretive measure as well as the relative importance of satisfying that criterion. A procedure will be presented in Sec. VII to achieve this result.

SUMMARY

The first step in formal assessment is to define explicitly what is desired in the way of performance from produced alternatives to complete a stated job. This means listing overall objectives or major performance criteria and insuring that the list is:

1. complete (i.e., contains all criteria which decision makers are able and willing to formulate and display);

2. mutually exclusive (i.e., contains criteria which neither encompass nor are encompassed by other criteria on the list);

3. of major significance (i.e., contains only highest-level criteria); and

4. free of worth interdependence (i.e., contains only worth-independent criteria).

Having established a list of overall performance objectives, the second step is to generate a hierarchical structure of successively more specific performance criteria. This involves breaking down or subdividing higher-level criteria into one or more lower-level criteria alleged to be included within the meaning thereof.

The third step is to select a single physical performance measure for each lowest-level performance criterion in the hierarchical structure. The purpose of selecting physical performance measures is to establish concrete connections between the hierarchical criterion structure (existing in the subjective minds of decision makers) and the outer world of physical alternatives.

However, merely establishing connections is not sufficient in itself to permit formal evaluation. Specific worth relationships must be mapped out between each lowest-level performance criterion and its related physical performance. This constitutes the fourth step. It is implemented by defining scoring functions which assign a unique worth score in points to every possible value of a physical performance measure. Scoring functions will be defined, either explicitly or implicitly, for every lowest-level criterion.

The fifth step is to combine worth scores assigned on the basis of separate performance criteria to arrive at a single overall index of worth. This is accomplished by defining a weighting function. An additive weighting function with constant trade-off weights will be adopted for this purpose. This requires that sets of sub-criteria located at every branch of the hierarchical tree contain members relatively independent in the worth sense. In addition, weights must be adjusted to reflect the differential interpretive quality of various performance measures.

This completes the outline of our assessment procedure. Step-by-step means of implementation will be presented in Section VII. Section VIII will illustrate the overall procedure with a complete example.

VII. IMPLEMENTING THE ASSESSMENT PROCEDURE

The procedure to be presented in this section is first of all intended to generate an assessment algorithm. This algorithm is supposed to encapsulate the worth notions of a particular decision maker (or group of decision makers) at a particular point in time with respect to a particular and clearly specified job. Once generated, the algorithm may then be applied to any feasible alternative produced to accomplish that job. Application of the algorithm to any one of the alternatives converts a description of that alternative, in terms of physical performance measures, into a single, overall index of that alternative's worth. It will be well to keep in mind the two-stage nature of the assessment procedure (i.e., first generate an assessment algorithm, and then apply the algorithm to generate a worth measure for each produced alternative). Otherwise, a confused interpretation may very likely result.

It is assumed that the following preliminary steps have been successfully completed prior to embarking upon the assessment process.

1. The job for which produced alternatives are being assessed has been adequately described.

2. From the job description a set of mandatory performance (and possibly resource) requirements has been extracted and recorded in physical terms.

3. At least two alternatives have been produced, one of which may be retaining the existing system, if one exists.

4. The performance and resource estimates associated with each produced alternative have been validated (i.e., investigated for accuracy and truthfulness).

5. These validated estimates have been checked against stipulated mandatory requirements, and at least two alternatives have been shown to be feasible. If fewer alternatives satisfy truly mandatory requirements, this is the signal to re-design existing alternatives and/or to produce new ones until at least two feasible proposals emerge. The need for at least one feasible alternative is obvious. Without any, there is no assessment problem to worry about. The need for at least two is not obvious at this time. However, the final steps in our assessment procedure (which have been shown experimentally to be critically important) require that at least two exist.

ESTABLISHING MAJOR OBJECTIVES

The process of establishing major objectives (highest-level performance criteria) was discussed in Section VI. Although it is difficult to reduce this process to a rigorous, step-by-step procedure, it is extremely important that great care be exercised. Non-rigorous procedure is recommended here to permit decision makers the widest possible creative latitude in deciding just what it is they want from an alternative. On the other hand, this is the time to discuss extensively and to resolve definitively such overriding issues as the following.

1. Whose interests are to be optimized, whose interests are to be minimally satisfied, and whose interests are to be ignored completely in choosing an alternative?

2. Concerning those whose interests are to be optimized, how can these interests be articulated in terms of clear overall objectives?

3. Concerning those whose interests are to be minimally satisfied, what will constitute minimum satisfaction of their needs?

Let us illustrate these issues and the operational consequences of their resoltuion by means of an example.

One crucial decision which the Department of Transportation must make prior to undertaking any formal assessment of alternatives concerns the precise manner in which they intend to balance off the frequently conflicting interests of users, operators, and the rest of society in the Northeast Corridor. This decision, in turn, depends upon DOT's view of its proper role as a governmental agency. Some alternative views are presented below, along with procedural implications.

Suppose DOT defined its role as follows:

1. to optimize the interests of users and society within the Northeast Corridor, making trade-offs, side payments, etc., between these two groups whenever a net gain in overall benefits might thereby be realized;

2. to provide manufacturers, contractors, operators, etc., the minimum amount of benefits (e.g., a minimum rate of return on investment) required to insure their participation in building and operating a transportation system; and

3. to ignore the interests of groups located outside of the Northeast Corridor.

Under these circumstances, the interests of both users and society would achieve explicit representation within the criterion hierarchy and within the list of mandatory performance requirements; the interests of manufacturers, contractors, and operators would appear <u>only</u> in terms of mandatory performance requirements; and the interests of non-corridor groups would achieve no explicit representation anywhere.

Alternatively, DOT might decide to optimize societal benefits, to minimally satisfy both users and operators, and to ignore all other considerations. Then the criterion hierarchy would reflect only societal interests. Consideration of user interests would be reduced to mandatory performance requirements, just like operators.

Our assessment procedure is not concerned with how DOT decides this issue. Either of the above alternatives (and many others) can be handled. However, the procedure requires that DOT make some decision on this issue and that the decision be made at the outset of formal assessment activities. Since trade-offs can only be made among criteria included in the hierarchy, it is essential that the scope of the hierarchy be decided in advance. In addition, since it is much easier to handle mandatory performance requirements than to effect trade-offs among criterion scores, it will save a great deal of time and effort if the criterion hierarchy is kept as small as possible.

Once the above issues have been resolved and a set of major objectives has been formulated, the only remaining step is to insure that the objectives are:

1. complete;
2. mutually exclusive;
3. highest-order; and
4. worth-independent.

These topics were discussed in Section VI.

The remainder of this section will concetrate almost exclusively upon the contents of the criterion hierarchy. Formulation of mandatory

performance requirements is left largely to engineering, economic, and/or political considerations.

GENERATING SUB-CRITERIA

As mentioned previously, it will be most helpful for implementing this and subsequent procedures if a master list of candidate performance criteria and performance measures has been compiled in advance. Although not necessary, experience has shown that reference to such a master list facilitates considerably the essentially creative process of filling out a criterion hierarchy and selecting performance measures. For purposes of discussion, it will be assumed that such a master list exists.

Beginning with one of the major or highest-level criteria, we ask what this means in the context of the stated job. To render the discussion concrete, let us assume that user and societal interests have been determined relevant candidates for optimization and that manufacturers, contractors, and operators are to be minimally satisfied. Then, there would be two major criteria:

1. user's interests, and
2. societal interests.

With reference to the job description, we might decide that the following sub-criteria are all intended by or subsumed under the major criterion "user interests":

1. passenger interests, and
2. freight interests.

Further subdivision of "passenger interests" might yield the following list:

1. travel time,
2. travel cost,
3. travel anxiety, and
4. travel comfort.

7. parking or returning private vehicle (time required);
8. unloading baggage (time required);
9. walking to and from other than the line-haul vehicle (time required);
10. waiting in queues (time required);
11. checking baggage (time required);
12. walking to and from line-haul vehicle (time required);
13. sitting passively on-board line-haul vehicle (time required);
14. claiming baggage (time required); and
15. retiring at night (earliest possible time of day).

The example partially developed in the preceding paragraphs could be carried further, but the general idea should by now be clear. One starts at the highest level of the hierarchy with one of the major performance criteria, asks himself what this means, defines one or more sub-criteria in response to this question, and then repeats the procedure with each of the defined sub-criteria. This process continues until it is decided that further subdivision is unwarranted. A physical performance measure is chosen, and that branch of the tree is considered filled out. A retreat is then made back up the tree to the first level containing incomplete branches. The process of successive subdivision is initiated at that point and carried out until another physical performance measure is defined. By so moving up and down the tree, an entire hierarchical structure may be generated. The final signal to stop occurs when no more incomplete branches exist (i.e., when physical performance measures have been attached to every branch of the tree).

Because the process just illustrated is recursive (i.e., because it involves successive reapplication of the same sequence of steps to move up and down the hierarchical tree), only the reiterated sequence of steps need be specified in any great detail to describe completely the entire process. A formal presentation of this reiterated sequence of steps follows immediately.

Step 1. Locate an incompleted branch on the hierarchial tree (i.e., any major criterion or sub-criterion without an attached physical performance measure). At the outset, incompleted branches will occur only at the top level of major performance criteria.

Step 2. With reference to the job description and to the master list, decide whether the criterion under scrutiny is to be further subdivided or interpreted directly by means of a physical performance measure. If it is to be further subdivided, proceed to Step 3. If a physical performance measure is to be selected for it, proceed to Step 5.

Step 3. Again, with reference to the job description and to the master list, subdivide the criterion under scrutiny into one or more sub-criteria. That is, decide what sub-criteria are intended by or logically subsumed beneath the criterion under scrutiny. Each of these now constitute new incompleted branches of the hierarchy.

Step 4. Choose any one of the sub-criteria defined in Step 3 as a starting point and return to Step 1.

Step 5. With reference to the job description and to the master list, select a physical performance measure judged relevant to the criterion under scrutiny.

Step 6. Move backwards up that particular branch of the hierarchy until the first level containing at least one incompleted branch is encountered. If this occurs at other than the top level of major criteria, choose the incompleted branch (any one of the incompleted branches if more than one exists), and return to Step 1 with this as a new starting point. If no incompleted branches are encountered until reaching the top level, proceed to Step 7.

Step 7. Inspect the top level of the hierarchical tree. If all major performance criteria have been completely "filled out" (i.e., if all branches starting at the top level have been completed), the process is over. A complete hierarchical structure has been constructed. However, if one or more incomplete branches remain, choose any one of those remaining as a starting point, and return to Step 1.

This completes the procedure. A tentative criterion structure has been created and given concrete interpretation by means of the physical measures attached, respectively, to each of the lowest-level criteria. Subsequent procedures will test this tentative structure for contamination by resource considerations and worth interdependent criteria. Also, the process of selecting physical performance measures (step 5) will be clarified.

IDENTIFYING RESOURCE CONSIDERATIONS

The reader will recall from Section IV that a distinction was made between worth considerations and resource considerations. It is at this point in the procedure that the distinction is implemented. Every lowest-level criterion and its associated performance measure must be reviewed for a possible resource interpretation.

A resource is any physical entity which is not desired directly in the context of the stated job, but which is indirectly desirable, since it may be converted through some physical process into an end-product which is directly desirable. Time and money are two examples of resources. Neither is desirable in and of itself (at least not in the transportation context), but both may be converted into desirable entities. A short procedure for identifying resource considerations appears below.

Step 1. Begin with the first physical performance measure at the base of the hierarchy.

Step 2. Ask which of the following two statements better describes the relationship between the physical measure and its related lowest-level criterion.

A. Satisfaction of the criterion is directly important in accomplishing the stated job, and the physical measure serves to indicate whether or not and to what extent that criterion has been satisfied.

B. Satisfaction of the criterion means nothing more in the context than conserving the associated physical measure so that it may later be converted into or exchanged for something else directly desirable.

If Statement A is selected as more descriptive than Statement B, move to the next performance measure, and repeat this question. Continue in this manner until all performance measures have been checked. On the other hand, if Statement B is more descriptive, set the physical measure aside for later consideration (Section IX will discuss combining worth scores with resource considerations). Delete the attached lowest-level criterion from the hierarchy. Then move to the next performance measure, and continue as before.

This completes the procedure. However, before moving on to the identification of worth interdependence, the reader's attention is redirected to the discussion in the above section. Recall that both "travel time" and "travel responsibility" were interpreted in terms of required times (in hours and minutes). Nevertheless, the meanings of these time measures were quite different. By applying the above procedure, "travel time" (and also "travel cost") would have been deleted from the hierarchy, while all of the responsibility items would have been retained. Why? Because passengers are concerned with "travel time" and "travel cost" only because expenditures thereof prevent expending the same resources on something else desirable. On the other hand, "time spent waiting in a queue" serves to measure how undesirable that waiting experience really is.

IDENTIFYING WORTH INTERDEPENDENCE

The section on Generating Sub-Criteria outlined a procedure for generating lower-level performance criteria intended by or included within the meaning of a higher-level criterion. This procedure was presented in step-by-step form. Step 3 in the procedure is the exact point at which a higher-level criterion is to be so subdivided. The question now is, what guidelines can be provided to aid in this process of subdivision?

Perhaps the best way to answer the question is to look at the final use to which subdivided criteria will be put. After an entire hierarchical worth structure has been formulated, decision makers will investigate first the set of major performance criteria and then each

set of sub-criteria. For every such set, they will determine the relative importance of each sub-criterion as a component of its related higher-level criterion. The determined relative importance of each sub-criterion will then be reflected by a numerical weight assigned thereto. Finally, these numerical weights will be used to transform intermediate point scores assigned to the sub-criteria (one score to each sub-criterion) into a single point score to be assigned to their related higher-level criterion.

Now it was pointed out in Section VI that use of an additive weighting function with constant weights is legitimate <u>only</u> when applied to performance criteria judged independent of one another in the worth sense. Therefore, whatever guidelines are developed to aid in the process of subdividing higher-level criteria should certainly include a means of identifying instances of substantial worth interdependence. Two specific questions are presented below to help distinguish worth-independent sub-criteria from those displaying substantial worth interdependence.

1. In comparing a candidate sub-criterion with its related higher-level criterion, which of the following statements better describes the apparent relationship between the two?

   (a) The sub-criterion is intended by, included within the meaning of, or an integral part of the higher-level criterion.

   (b) The sub-criterion is one alternative means of satisfying the higher-level criterion and important only insofar as it contributes thereto.

2. In comparing one candidate sub-criterion with another sub-criterion already judged as appropriately included within the same set, which of the following statements better describes the apparent relationship between the two?

(a) Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would not be influenced by the degree of satisfaction already obtained on each.

(b) Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would depend markedly on the degree of satisfaction already obtained on each.

In order to qualify for final inclusion in the hierarchical structure, every candidate sub-criterion must receive an "a" answer to both of the above questions.

A specific, step-by-step procedure incorporating the above pair of questions appears below. It is intended that a first pass be made at creating a tentative criterion structure by means of the procedure presented in Generating Sub-Criteria. Then, this procedure may be applied to the candidate sub-criteria generated thereby. An alternative approach would be to perform this testing procedure every time a higher-level criterion is subdivided into a set of sub-criteria (i.e., after Step 3 in the procedure on Generating Sub-Criteria). Either approach would work; however, the step-by-step procedures have been written under the assumption that they will be performed sequentially rather than concurrently.

Step 1. Begin with any set of candidate sub-criteria previously generated in filling out the hierarchical structure (see Generating Sub-Criteria, Step 3, for the exact point at which a set of candidate sub-criteria is generated).

Step 2. Arrange them in a sequence. It makes no difference how they are arranged--any arbitrary sequence will suffice.

Step 3. Compare the first sub-criterion in the sequence with the higher-level criterion of which all of the sub-criteria are alleged to be component parts. Ask which of the following two statements better

describes the relationship between the candidate sub-criterion under scrutiny and its related higher-level criterion.

A. The sub-criterion is intended by, included within the meaning of, or an integral part of the higher-level criterion.

B. The sub-criterion is one alternative means of satisfying the higher-level criterion and important only insofar as it contributes thereto.

If statement A is selected as more descriptive than statement B, move to the next sub-criterion in the sequence, and repeat the same question regarding its relationship to the higher-level criterion. Continue in this manner until the entire sequence has been exhausted; then proceed to Step 4. On the other hand, if statement B is selected as more descriptive than statment A, the sub-criterion under scrutiny does not properly belong in the set. Delete this sub-criterion from the set, lay it aside temporarily, and reconsider it later. (Note: Suggested procedures for handling deleted sub-criteria are discussed later in this paper). Move to the next candidate sub-criterion in the sequence and repeat the same question, continuing in this manner until the entire sequence has been exhausted.

Step 4. Select another set of candidate sub-criteria as yet unchecked for worth interdependence, and return to Step 2. If all sets of sub-criteria have been checked, proceed to Step 5.

Step 5. At this point, the entire hierarchical worth structure has been tested (at least partially) for worth interdependence. Quite possibly, some candidate sub-criteria have been deleted and set aside pending subsequent reconsideration. However, it will be useful to check the remaining structure to insure that all sub-criteria are really worth-independent. This can be accomplished by repeating Steps 1 through 4 on the entire hierarchy, but with a new question substituted for the old question in Step 3. A revised form of Step 3 is presented below to facilitate this "second pass" at testing the hierarchy.

Next, with reference to the master list we might discover a fifth sub-criterion which is not suggested directly by the job description, but which we consider to be a definite component of "passenger interests." This fifth item might be "travel responsibility," which denotes the various ways in which passengers must assume personal responsibility for managing their own activities along the way.

If we feel that this more or less exhausts the intended meaning of "passenger interests," we can proceed to process each of the five sub-criteria just generated in a similar manner.

Beginning with "travel time," we ask ourselves the same question. What does this mean in the context of the stated job? At this point, we may decide that further subdivision is unnecessary. An obvious performance measure suggests itself--namely, the "time required in minutes to make the trip."

Returning to the second sub-criterion, "travel cost," we can select a performance measure straightaway. "Total dollar expenditures required to finance the entire trip" would seem to encapsulate the meaning of this sub-criterion quite well.

Once again we return to the next-higher branch on the tree. Suppose we select the fifth sub-criterion, "travel responsibility." We might decide to subdivide this into the following fifteen items, each one representing a distinct phase in the origin-to-destination trip. Furthermore, we might decide to interpret each item directly in terms of the physical measures shown below:

1. responsibility to arise early enough in the morning (latest possible time of day);

2. responsibility to allow extra or "pad" time for unforeseen contingencies (time allowed in minutes);

3. responsibility for searching or renting a private access vehicle (time required);

4. loading baggage (time required);

5. operating private vehicle (time required);

6. maintaining private vehicle (time required);

Step 3 (revised). Compare every possible pair of sub-criteria in the sequence. (Note: If there are N sub-criteria in the sequence, there are $\frac{1}{2}N^2 - \frac{1}{2}N$ such pair-wise comparisons to be made.) Ask which of the following two statements better describes each pair-wise relationship.

A. Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would not be influenced by the degree of satisfaction already obtained on each.

B. Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would depend markedly on the degree of satisfaction already obtained on each.

If statement A is selected as more descriptive than statement B, move to the next pair of sub-criteria, and repeat the same question. Continue in this manner until all of the $\frac{1}{2}N^2 - \frac{1}{2}N$ pair-wise comparisons have been made; then proceed to Step 4. On the other hand, if statement B is selected as more descriptive than statement A, at least one of the sub-criteria in the pair-wise comparison does not properly belong in the set. Move to the next pair of sub-criteria, and repeat the same question. Continue in this manner until all pair-wise comparisons have been made. Then, by inspecting pairs which contain at least one improper member, delete and set aside those sub-criteria which do not belong in the set pending subsequent reconsideration.

This completes the identification procedure.

SELECTING PHYSICAL PERFORMANCE MEASURES

Let us now investigate the task of selecting physical performance measures. After accomplishing sufficient conceptual refinement through successive subdivisions of higher-level criteria into sets of lower-level criteria, a single performance measure must be chosen to interpret concretely each of the lowest-level criteria in the generated hierarchical structure. In essence, our problem is to select for each lowest-level criterion some physically measurable attribute which is perceived as

embodying or providing a concrete interpretation of that criterion. Thus, if one lowest-level criterion were "waiting time," then "time waiting in a passenger terminal on a typical trip" might provide a suitable measure. If this time were highly variable over trips, then "average time waiting in a passenger terminal" might provide a better measure of the unpleasantness which could be anticipated. If some trips involve extremely large waiting times, then the average might not be a good measure either. Possibly the maximum time would be better still.

From the preceding illustrative discussion, the reader may be somewhat disturbed to see that more than one physical performance measure may be applicable to any given lowest-level performance criterion. Furthermore, where more than one performance measure appears applicable, it may not always be obvious which one to choose. In short, judgment on the part of the decision maker must again be exercised to select an appropriate measure just as it was in generating sub-criteria.

One important factor to consider in selecting performance measures is their practical feasibility. Only measures for which complete and timely data may be obtained are feasible.

Another factor is the question of their order or degree of generality. An example of an extremely high-order measure would be the "total time required to transport a carload of strawberries from New York to Boston." This would reflect numerous lower-level measures of intermediate point-to-point transportation times.

An example of a moderately high-order measure would be "line-haul delay time" contributed by various sources of delay along the line-haul portions of the complete trip.

Contrast each of these two examples with "waiting time at the baggage claim area." This latter measure is extremely low-order. It is difficult to decompose this into more elementary component measures.

Now the order of a performance measure is important for two reasons. First, high-order measures are generally more relevant for assessment than are lower-order measures, unless division among component lower-order measures serves to convey different evaluative significance in the stated job context. Consequently, an effort should be made to select and/or concoct high-order measures whenever possible. Second,

creation of high-order measures out of lower-order component measures can often be used to retrieve sub-criteria temporarily deleted from the criterion structure due to worth interdependence. Such deleted sub-criteria can be replaced by a single higher-level criterion, and a single high-order performance measure can be selected to go with it. Thus, "total line-haul delay" for freight could replace the delay times contributed by various sources. (Note: This device is utilized for freight, but not for passenger delays because cargo does not care how or why it is delayed, while people react differently under different circumstances of delay.)

In summary, what guidelines can now be provided for the selection of appropriate physical performance measures? Five guidelines are suggested.

1. Consult the master list to obtain a set of candidate measures.

2. Augment this set by inventing any additional measure not contained in the master list, but which seems appropriate in the context of the lowest-level criterion under consideration and the stated job.

3. Check candidate measures for practical feasibility (i.e., to insure that all data included in the measure can be conveniently and promptly gathered).

4. Attempt to combine candidate measures into higher-order measures, where possible and appropriate.

5. On an intuitive basis, select the seemingly most appropriate and highest-order of the practically feasible candidate measures.

A specific step-by-step procedure incorporating the above five guidelines appears below. It is intended that this procedure be implemented concurrently with the generating procedure outlined in Generating Sub-Criteria.

Step 1. Begin with any one of the lowest-level performance criteria occurring at the base of the previously generated hierarchical structure.

Step 2. Consult the master list of performance criteria and physical performance measures. Looking only at the physical measures

contained in the master list, identify those which are perceived as significantly related to the lowest-level criterion under consideration. This may be done by asking the following question about the relationship perceived to exist between the criterion and every physical performance measure on the master list.

> Would changes in the state or numerical value of the performance measure be capable of bringing about either significant increases or significant decreases in the extent to which the lowest-level criterion under consideration is satisfied?

If the answer to the above question is yes, then a significant relationship is said to exist between the lowest-level criterion and the physical performance measure. If the answer is no, then no such relationship is perceived.

Step 3. Add to the set of physical measures drawn from the master list and perceived as significantly related to the lowest-level criterion any additional measures which can be thought of and which also seem related. In this manner, decision makers can supplement the master list with their own imagination and experience.

Step 4. Looking now at all candidate measures generated by Steps 2 and 3, check to see whether each is practically feasible. That is, insure that all data necessary to form each measure can be conveniently and promptly gathered. Delete any candidate measures which are discovered to be practically infeasible.

Step 5. Inspect the residue of feasible candidate measures remaining after Step 4. Either choose one straightaway (by intuitive judgment) as the most appropriate single measure by which to interpret the lowest-level criterion or, if none of the feasible candidates seem really appropriate, attempt to construct a higher-order physical measure out of two or more individual measures.

Step 6. Proceed to another lowest-level criterion in the hierarchical structure, and repeat Steps 2 through 5. Continue in this manner until all lowest-level criteria have been assigned a corresponding physical performance measure. (Note: It may be that no performance

measures can be found for some lowest-level criteria. When this occurs, a direct worth estimate must be made by the decision maker without the aid of either a physical measure or a scoring function. Scoring functions are considered to be implicit in the mind of the decision maker under these circumstances.)

MAPPING OUT SCORING FUNCTIONS

The next task is to formulate scoring functions by which each lowest-level performance criterion may be linked to its assigned measure of physical performance. Once formulated, these scoring functions may be used to convert measured physical performance into equivalent worth scores, and these worth scores may then be combined via weights into a single numerical index indicating the overall worth of a proposed alternative.

The scoring procedure itself will be broken down into two major phases. The first phase will contain an ordered sequence of questions designed to determine the general nature and shape of whatever scoring function is to be formulated. The nature and shape of each scoring function will be inferred from answers to the following questions.

1. Is the physical performance measure to be scored discrete or continuous?

2. If discrete, how many measurement categories are contained in the physical performance scale; is there any inherent order or sequence built into this scale; and are there any qualitative distinctions to be made concerning observations within each measurement category?

3. If continuous, is the physical performance scale bounded from above and/or from below?

4. If bounded, where do the boundaries of the physical performance scale fall?

5. With which points on the physical performance scale are zero and one hundred percent satisfaction of the related lowest-level performance criterion associated, respectively.

6. Does satisfaction increase or decrease with increases in measured performance?

7. Does the rate at which satisfaction increases or decreases with increases in measured performance ever change, or does it remain constant over the entire range of the physical performance scale?

8. If the above rate changes, does it always increase, or does it always decrease, or does it both increase and decrease over selected intervals within the range of the physical performance scale?

The second phase will contain a step-by-step procedure designed to select a specific scoring function of the general nature and shape indicated in the first phase. Actually, two alternative procedures will be presented to implement this second phase--one involving visual and graphic methods, and the other involving numerical methods. The choice between these two alternative procedures will be left up to the discretion of decision makers.

The ordered sequence of questions designed to implement Phase I is displayed below. All Phase II procedures are referenced by these questions and appear in the Appendices at the end of this paper. (Note: All functions are assumed positive in the discussion that follows.)

Step 1. Consider the scale of the physical performance measure. Is it continuous or is it discrete? If discrete, proceed to Step 2. If continuous, proceed to Step 7.

Step 2. Is the discrete scale purely discrete or is it a hybrid, containing continuous aspects as well as discrete aspects? If purely discrete, proceed to Step 3. If hybrid, treat it as if it were continuous and proceed to Step 7.

Step 3. How many categories or levels are contained within the discrete scale identified in Step 2? If two, proceed to Step 4. If three, four or five, proceed to Step 5. If more than five, proceed to Step 6.

Step 4. If the discrete, two-level scale identified in Step 3 is merely a case of presence or absence of some desirable attribute, proceed to scoring procedure 1 in Appendix A. If presence of the desirable attribute is to be qualified by an additional measure of relative worth, proceed to scoring procedure 2 in Appendix B.

Step 5. Is the discrete scale identified in Step 3 strictly nominal or is it ordered? If strictly nominal, proceed to scoring procedure 3 in Appendix C. If ordered, proceed to scoring procedure 4 in Appendix D.

<u>Step 6</u>. Is the discrete scale identified in Step 3 strictly nominal or is it ordered? If strictly nominal, proceed to scoring procedure 5 in Appendix E. If ordered, treat the scale as if it were continuous, and proceed to Step 7.

<u>Step 7</u>. Does the continuous scale identified in Step 1, Step 2, or Step 6 possess a logical lower bound? If yes, proceed to Step 10. If no, proceed to Step 8.

<u>Step 8</u>. It is very unlikely that a performance measure will have been selected whose scale is unbounded from below (i.e., where negative observations are possible and may range all the way to negative infinity). Therefore, ask once again whether the scale under scrutiny possesses a logical lower bound. If the answer is now yes, proceed to Step 10. If the answer is still no, look for a logical upper bound. If the scale possesses no logical upper bound either, the performance measure must be rejected. The scoring procedures presented herein are not equipped to handle doubly unbounded performance scales. Choose a new performance measure, and return to Step 1. However, if the scale does possess a logical upper bound, proceed to Step 9.

<u>Step 9</u>. Transform the scale identified in Step 8 by multiplying every number contained therein by minus one. This transformed scale will now possess a logical lower bound, but no logical upper bound. Proceed to Step 10, but keep in mind that the new transformed scale is just the reverse of the original scale. Consequently, all subsequent questions about the transformed scale must be answered with this reversed aspect in mind.

<u>Step 10</u>. Does the logical lower bound fall exactly at zero? If yes, proceed to Step 12. If no, proceed to Step 11.

<u>Step 11</u>. Identify the numerical value of the logical lower bound. Transform the scale by subtracting this number from every number contained in the scale. Keep this transformation in mind, and remember that all subsequent questions will refer to the transformed scale. Proceed to Step 12.

Step 12. Does the scale possess a logical upper bound? If yes, proceed to Step 13. If no, proceed to Step 24.

Step 13. It has been determined that the performance scale is bounded from below by zero and from above by some finite positive number. What is the direction of the preference relationship? If direct (i.e., more is better), proceed to Step 14. If reverse (i.e., less is better), proceed to Step 19.

Step 14. Now fit the end-points of the worth scale to the logical lower and upper bounds of the performance scale. Assign zero worth points to zero performance and one worth point to the logical upper bound of the performance scale. Proceed to Step 15.

Step 15. Is the direct preference relationship identified in Step 13 uniform over the entire logical range of the performance scale? If yes, proceed to Step 16. If no, sketch the approximate shape of the preference relationship, and proceed directly to scoring procedure 20 in Appendix T.

Step 16. Does the direct preference relationship identified in Step 15 maintain a constant rate of change of worth, or does it display a variable rate of change (i.e., either accelerating, decelerating, or both in sequence)? If constant, proceed to scoring procedure 6 in Appendix F. If variable, proceed to Step 17.

Step 17. Does the variable rate of change of worth identified in Step 16 display uniform acceleration, uniform deceleration, or first one and then the other? If uniform acceleration, proceed to scoring procedure 8 in Appendix H. If first one and then the other, proceed to Step 18.

Step 18. Does the variable rate of change identified in Step 17 start by accelerating and then end by decelerating, or does it start by decelerating and then end by accelerating? If the former, proceed to scoring procedure 9 in Appendix I. If the latter, proceed to scoring procedure 10 in Appendix J.

Step 19. Now fit the end-points of the worth scale to the logical lower and upper bounds of the performance scale. Assign zero worth points to the logical upper bound of the performance scale, and assign one worth point to zero performance. Proceed to Step 20.

Step 20. Is the reverse preference relationship identified in Step 13 uniform over the entire logical range of the performance scale? If yes, proceed to Step 21. If no, sketch the approximate shape of the preference relationship, and proceed directly to scoring procedure 20 in Appendix T.

Step 21. Does the reverse preference relationship identified in Step 20 maintain a constant rate of change of worth, or does it display a variable rate of change (i.e., either accelerating, decelerating, or both in sequence)? If constant, proceed to scoring procedure 11 in Appendix K. If variable, proceed to Step 22.

Step 22. Does the variable rate of change of worth identified in Step 21 display uniform acceleration, uniform deceleration, or first one and then the other? If uniform acceleration, proceed to scoring procedure 12 in Appendix L. If uniform deceleration, proceed to scoring procedure 13 in Appendix M. If first one and then the other, proceed to Step 23.

Step 23. Does the variable rate of change identified in Step 22 start by accelerating and then end by decelerating, or does it start by decelerating and then end by accelerating? If the former, proceed to scoring procedure 12 in Appendix N. If the latter, proceed to scoring procedure 15 in Appendix O.

Step 24. It has been determined that the performance scale is bound from below by zero, but that the scale possesses no logical upper bound. What is the direction of the preference relationship? If direct (i.e., more is better), proceed to Step 25. If reverse (i.e., less is better), proceed to Step 28.

Step 25. Now fit the end-points of the worth scale to the performance scale. Assign zero worth points to zero performance and one worth point to infinite performance. Proceed to Step 26.

<u>Step 26</u>. Is the direct preference relationship identified in Step 24 uniform over the entire logical range of the performance scale? If yes, proceed to Step 27. If no, sketch the approximate shape of the preference relationship, and proceed directly to scoring procedure 20 in Appendix T.

<u>Step 27</u>. The following facts have been ascertained concerning the nature and shape of the scoring function for this performance measure.

1. The worth scale is bounded between zero and one (by convention).

2. The physical performance scale is bounded from below by zero, but it possesses no logical upper bound.

3. The preference relationship is uniformly direct over the entire range of the performance scale.

From these three facts, we must conclude that both a constant rate of change of worth and a uniformly accelerating rate of change of worth are logically impossible. The only remaining possibilities discussed herein are uniform deceleration or initial acceleration followed by deceleration. If uniform deceleration, proceed to scoring procedure 16 in Appendix P. If initial acceleration followed by deceleration, proceed to scoring procedure 17 in Appendix Q.

<u>Step 28</u>. Now fit the end-points of the worth scale to the performance scale. Assign one worth point to zero performance and zero worth points to infinite performance. Proceed to Step 29.

<u>Step 29</u>. Is the reverse preference relationship identified in Step 24 uniform over the entire logical range of the performance scale? If yes, proceed to Step 30. If no, sketch the approximate shape of the preference relationship, and proceed directly to scoring procedure in Appendix T.

<u>Step 30</u>. The following facts have been ascertained concerning the nature and shape of the scoring function for this performance measure.

1. The worth scale is bounded between zero and one (by convention).

2. The physical performance scale is bounded from below by zero, but it possess no logical upper bound.

3. The preference relationship is uniformly reverse over the entire range of the performance scale.

From these facts, we must conclude that both a constant rate of change of worth and a uniformly accelerating rate of change of worth are logically impossible. The only remaining possibilities discussed here are uniform deceleration or initial acceleration followed by deceleration. If uniform deceleration, proceed to scoring procedure 18 in Appendix R. If initial acceleration followed by deceleration, proceed to scoring procedure 19 in Appendix S.

This completes the ordered sequence of questions designed to determine the general nature and shape of the scoring function.

ASSIGNING WEIGHTS

The weight-setting procedure to be developed herein is divided into two sequential phases. In the first phase, an individual decision maker attempts to produce his own numerical weights corresponding to each of the sub-criteria contained in some specified set of sub-criteria appearing in the hierarchical structure. In the second phase, individual weight sets assigned by separate decision makers are compared, and lack of consensus among decision makers (if there is more than one) is resolved by an averaging technique.

The first phase of the procedure involves two major operations.

1. All sub-criteria subsumed under a given higher-level criterion are ranked in order of ascending perceived importance.

2. Then, starting with the most important pair of sub-criteria appearing at the head of the list, successive pair-wise comparisons are made between contiguous sub-criteria, and decision makers are asked to indicate in terms of a ratio the degree of perceived relative importance of the two. Stated alternatively, decision makers are asked to indicate the rate at which they would be willing to accept reduced satisfaction of one sub-criterion in return for increased satisfaction of the other.

A step-by-step procedure to implement this first phase follows immediately. The resulting individual weights generated by this procedure are all positive, they sum to one, and they are interpretable in accordance with the weighting conventions stipulated in Section VI. However, one word of warning seems appropriate. Although this procedure guarantees that the resultant weights will possess certain desirable logical properties (i.e., consistency, transitivity, and preservation of the preselected importance ratios), the validity of the weights themselves still remains the responsibility of informed judgment on the part of decision makers. Neither this procedure nor any other procedure based solely on logical considerations can guarantee their validity. Only clearly articulated judgment can ever provide that. Additional procedures designed to aid in this validation process will be presented at the end of Section VII.

Step 1. Begin with any set of sub-criteria subsumed under a higher-level performance criterion.

Step 2. List these sub-criteria in approximate order of relative importance, starting with the most important sub-criterion at the top of the list and the least important sub-criterion at the bottom. It is not necessary to have the sub-criteria perfectly ranked or ordered on this first pass, since subsequent operations will be performed to guarantee complete ordering.

Step 3. Compare the first two-sub-criteria on the list.

a. If the first sub-criterion is deemed relatively more important than the second, proceed directly to Step 4.

b. If both sub-criteria are deemed roughly equal in importance, proceed directly to Step 4.

c. If the second sub-criterion is deemed relatively more important than the first, invert their positions on the list (i.e., place the first sub-criterion where the second used to be on the list, and vice-versa), and then proceed to Step 4.

Step 4. Compare the lower-ranked sub-criterion from Step 3 with the next sub-criterion on the list. Repeat the comparisons and stipulated operations in Step 3 on this new pair of sub-criteria. Continue in this manner all the way down to the end of the list until pair-wise comparisons have been made between all contiguous criteria.

Step.5. After the list has been completely exhausted, go back and determine whether any inversions (position changes) occurred.

a. If none occurred, proceed directly to Step 6.

b. If one or more occurred, return to the head of the list, and repeat the entire procedure described in Steps 3 and 4.

Step 6. Eventually, the list will become so arranged that successive pair-wise comparisons will generate no inversions. It may require several passes to achieve this result, but it will occur in the end (assuming that the decision maker's notions of relative importance among sub-criteria are both consistent and transitive). When the list has achieved an arrangement wherein no inversions occur, it will then reflect the decision maker's judgments of relative importance in terms of direction, but not yet in terms of magnitude. Relative magnitudes are determined by subsequent steps.

Step 7. Take the first sub-criterion on the rearranged list, and assign to it the number 1.0 or one hundred percent.

Step 8. Compare the second sub-criterion with the first, and assess their relative importance in terms of a ratio or fraction. That is, if satisfying the second sub-criterion seems only one-half as important as satisfying the first, assign the fraction 1/2 or its decimal equivalent .5 to the second sub-criterion. In like manner, fractions such as 3/4, 9/10, etc. or their decimal equivalents might equally well have been assigned. (Note: It may be difficult to set weights when the question is phrased in the above manner. An alternative form of the same question would be, "At what rate would reduced satisfaction of the first sub-criterion be acceptable in return for increased satisfaction of the second

so as to maintain the same overall worth considering satisfaction of both sub-criteria jointly?" The answer to this question, expressed in the form of a ratio, may then be assigned as before to the second sub-criterion.)

Step 9. Compare the second and third sub-criteria, assess their relative importance or trade-off rate in terms of either a fraction or a ratio, multiply the number assigned to the second sub-criterion by this fraction or ratio, and assign the resultant product to the third sub-criterion. For example, assuming that the second sub-criterion were assessed as being 1/2 as important as the first, while the third were assessed as being 9/10 as important as the second, the appropriate computation would be $1/2 \times 9/10$, and the number 9/20 would be assigned to the third sub-criterion.

Step 10. Repeat the above procedure for all successive pair-wise comparisons until the list of sub-criteria has been completely exhausted. Then, each sub-criterion will have been assigned a number equal to the product of its importance relative to the next higher sub-criterion times the number previously assigned to the next higher sub-criterion.

Step 11. Add the numbers assigned to all sub-criteria on the list, and then divide each one by the computed sum. This will serve to convert relative importance ratios into normalized weights. Each weight will be positive, and the whole set will add to one. In addition, the relative importance ratios will be preserved in the ratios of any pair of weights.

This completes the procedure.

Now there may not always be complete agreement among separate decision makers concerning the proper collection of weights to be attached to any set of sub-criteria. In fact, numerical differences, and perhaps even rank-order differences, are to be expected among separate decision makers -- particularly if they set weights without first consulting one another. This lack of consensus would seem quite healthy, in the writer's opinion, and should be encouraged

rather than discouraged. Unless any single decision maker is willing to claim that his weights are precisely correct and, therefore, that anybody who disagrees with him is necessarily wrong, then some method for combining group opinion would seem appropriate.

One way of combining group opinion would be to subject differences of opinion to open discussion in hopes of achieving greater consensus. This would be a particularly effective remedy for those situations where some decision makers possess greater knowledge and experience than others. By open discussion, the less knowledgeable and less experienced decision makers could benefit from their better endowed compatriots and thereby gain a sounder basis for assessment.

However, open discussion would <u>not</u> be effective against genuine differences of opinion held by equally knowledgeable and equally experienced decision makers. <u>Nor</u> would it be effective against whatever differences remain after open discussion has enlightened those decision makers who did not possess initially the same knowledge and experience as others, but who altered their opinions somewhat in the face of ensuing discussions. Some sort of compromise procedure would seem appropriate in these two instances.

One way of achieving a compromise would be by averaging individual weights across separate decision makers. That is, to each sub-criterion in a particular set, separate decision makers would assign their own individual weights. Then, an average weight would be computed for each sub-criterion by adding the weights assigned by separate decision makers and dividing the total by the number of decision makers. It can be shown that, if this averaging procedure is applied to each sub-criterion in a set, then the computed average weights assigned to each of the sub-criteria will sum to one. In addition, the resultant average weights would reflect group opinion instead of one single individual's opinion.

In actual practice, both of the above procedures would seem appropriate, if carried out in sequence. First, a group of decision makers would meet to discuss the relative importance of sub-criteria in some designated set. By open discussion, all decision makers would be accorded a similar basis for formulating their own individual

opinions. Then, each decision maker would reflect his individual opinion in a set of numerical weights (generated via the step-by-step procedure just presented). Finally, remaining differences of opinion would be handled by averaging over individual weight assignments to arrive at a final set of group weights for each of the sub-criteria. In this manner, spurious differences of opinion arising from differences in knowledge and experience would be minimized while genuine differences of opinion arising from genuinely different views of the situation would be adequately reflected in the final average weights.

A step-by-step procedure to implement the second phase of the weight-setting process, designed to average out remaining differences of opinion, is presented below.

Step 1. Collect whatever individual weights have been assigned by separate decision makers to a set of sub-criteria in the hierarchical structure.

Step 2. Suppose that there are N separate decision makers and M sub-criteria in the set to which individual weights have been attached. (Note: Both N and M are assumed to be greater than one. If N = 1, there would be no problem of lack of consensus. If M = 1, there would be only one sub-criterion in the set, and it would therefore have to receive a full weight of 1.0.)

Step 3. Lay out the individual weights assigned by separate decision makers in N parallel columns of M weights each. The resulting rectangular array may be thought of as a matrix with M rows and N columns.

Step 4. Compute and record the sum of the weights appearing in each of the M rows of the above matrix. (Note: If it is considered desirable to weight some opinions more heavily than others, compute an appropriately weighted sum.)

Step 5. Divide each computed row sum by N. This gives an average weight, averaged across the N separate decision makers, for each of the M sub-criteria. (Note: The M average weights must add

to one -- except, perhaps, for small rounding errors. If they do not add to one, check the computations for algebraic errors.)

This completes the procedure.

ADJUSTING THE WEIGHTS

The last step in formulating an assessment algorithm is to adjust the weights to reflect differential interpretive quality among the physical performance measures. A step-by-step procedure to accomplish this is presented below.

Step 1. Compute the "effective" weight associated with each lowest-level performance criterion. That is, identify the chain of weights linking each lowest-level criterion to the apex of the hierarchy, and compute the product of all weights in this chain. Then, each of the "effective" weights associated, respectively, with one of the lowest-level criteria will be positive, and they will sum to one.

Step 2. Now consider the relationship between each lowest-level criterion and its associated physical performance measure. Recalling the scoring function which has been defined for each of these linked pairs, assess the extent to which the performance measure serves to interpret, through its scoring function, the intended meaning of the lowest-level criterion. Assess its interpretive quality on a percentage scale, where zero means that the performance measure bears no relation at all to the performance criterion, and one hundred percent means that the performance measure interprets perfectly the intended meaning of that criterion.

Step 3. Assign percentage numbers to each linked pair at the base of the criterion structure.

Step 4. Multiply each "effective" weight by the corresponding percentage number assigned in Step 3.

Step 5. Add the products computed in Step 4.

Step 6. Divide each product computed in Step 4 by the sum computed in Step 5. The result is a set of "adjusted effective" weights.

This completes the procedure.

TESTING THE ENTIRE ASSESSMENT ALGORITHM

The process of formulating an assessment algorithm was completed in Adjusting the Weights, above. Now it is time to test the algorithm against real alternatives. The importance of performing this part of the task cannot be overemphasized. On the basis of an experiment (reported in Appendix U), it was determined that the efficacy of the entire assessment procedure depends critically upon performing the following steps. Incidentally, it is these steps that require at least two feasible alternatives prior to undertaking the task of formal assessment.

Step 1. Select one of the feasible alternatives.

Step 2. Select one of the performance measures in terms of which that alternative has been described.

Step 3. Referring to the scoring function associated with that performance measure, convert measured performance into an equivalent worth score.

Step 4. Multiply the equivalent worth score computed in Step 3 by the associated "adjusted effective" weight computed in Adjusting the Weights, above.

Step 5. Repeat Steps 2 through 5 for all performance measures.

Step 6. Add the products computed in Step 5. The resulting sum constitutes an index of the selected alternative's overall worth.

Step 7. Lay out all of the data computed in Steps 1 through 6 in some form convenient for comparisons across alternatives. A suggested format will be presented in the various exhibits of Section VIII.

Step 8. Select any subset of performance measures (possibly the entire set). On the basis of the selected subset, rank the alternatives

in order of perceived overall worth. However, do not consult any of the scores or weights associated with this subset of performance measures. Assign ranks solely on the basis of intuitive judgment based only on the physical measures themselves.

Step 9. Compute the partial worth score (possibly the total worth score, if all performance measures were considered) associated with each alternative. Compare the rank-order of these computed partial worth scores against the subjectively assigned ranks generated in Step 8. If there is complete agreement, proceed to Step 11. Otherwise, proceed to Step 10.

Step 10. Some disagreement has arisen between subjective and computed ranks. Common reasons for this are listed below:

1. incomplete list of criteria;
2. criteria contaminated with resource considerations;
3. criteria contaminated with interdependencies;
4. incorrect scoring functions;
5. measured performance lies beyond region of reasonable trade-offs as intended by assigned scoring functions and weights; or
6. incorrect weights and/or adjusting factors.

Attempt to diagnose the difficulty, and return to whichever portions of the complete assessment procedure require repair.

Step 11. If all reasonable tests have been made, stop. Otherwise, return to Step 8.

This completes the procedure.

VIII. TWO EXAMPLES

Undoubtedly, the best way to illustrate the foregoing procedure would be to apply it fully to the entire set of alternative transportation systems currently being considered for the Northeast Corridor. But this is a Herculean task. Not only is it impossible to implement, since a complete assessment has not yet been undertaken, it is also unnecessary. A partial example of its application to the Northeast Corridor problem should suffice to illustrate how our procedure can be made to work in this context. On the other hand, a partial example, although relevant to the transportation problem, would not suffice to illustrate every step of our procedure from start to finish. Consequently, we shall strike a compromise between relevance and completeness.

One example has been prepared concerning six alternative ways in which an average, middle-income businessman might travel from Washington to New York. In all six cases, the complete origin-to-destination trip is assessed. However, this constitutes only a partial example of the assessment procedure, since the impact of the six alternatives on freight, operator, societal, and other kinds of passenger interests is omitted. Only the average, middle-income businessman's point of view is considered.

This example was worked out by a team of RAND personnel over a period of several months. The results have been written up by T. F. Kirkwood and appear in a separate paper.*

Another example, although unrelated to transportation, has been included in this section. Since it is complete and self-contained, it will serve to illustrate our assessment procedure from start to finish.

BACKGROUND

One of the writer's acquaintances, a graduate student of Massachusetts Institute of Technology, became interested in the assessment procedure when he was faced with securing employment directly following

---

*See RM-5869-DOT, Measurement and Evaluation of Transportation Effectiveness by Frederick S. Pardee et al., especially Sec. D-X, "Passenger Trip Analysis," The RAND Corporation, 1969 (forthcoming).

graduation. He had already solicited several job offers and, on the basis of preliminary analysis, he had reduced these to a set of four feasible and reasonable alternatives. It was at this point that he undertook the task of formal assessment.

After reading completely a description of the procedure and obtaining clarification on various details from the writer, he set out to generate a criterion hierarchy, to establish weights, to define scoring functions, to adjust the weights, to assess the four alternative job offers, and, finally, to make a terminal decision. His progress through these sequential steps will be reported below.

THE CRITERION HIERARCHY

It would require too much space to present a complete historical record of this individual's progress through the various procedures involved in generating a criterion hierarchy, purging it of worth-interdependent members, and selecting physical performance measures. He made at least four separate passes at creating and revising a hierarchy over a period of several weeks time. What will be presented instead is the end state of this process. The hierarchy of worth-independent criteria and associated performance measures which he finally selected as providing a satisfactory description of his job objectives is described below.

Four major objectives or highest-level performance criteria were defined:

1. monetary compensation;
2. geographical location;
3. travel requirements; and
4. nature of work.

Monetary compensation was broken down to include:

1. immediate compensation; and
2. future compensation.

Immediate compensation was further subdivided to include:

1. starting salary; and
2. fringe benefits, which included -
   (a) insurance benefits; and
   (b) retirement benefits.

Future compensation was subdivided to include:

1. anticipated salary in three years; and
2. anticipated salary in five years.

His second major objective, geographical location, was broken down to include:

1. proximity to relatives;
2. degree of urbanity associated with the location; and
3. climate.

His third major objective, travel requirements, was broken down to include:

1. daily commuting requirements to and from the place of work; and
2. extended trips.

Extended trips was further subdivided to reflect:

1. proportion of time away from home; and
2. duration of extended trips.

His fourth major objective, nature of work, was broken down to include:

1. immediate training requirements; and
2. continuing aspects.

Continuing aspects of the work were further subdivided to include:

1. personal interest in the technical content of the job;

2. degree of variety implicit in the job; and

3. amount of training in management skills relizable from the job.

The above hierarchy contained fifteen lowest-level criteria, each one of which was interpreted by defining a single performance measure. These fifteen lowest-level criteria and their associated performance measures were as follows:

1. starting salary -- locally adjusted after-tax annual dollars;*

2. insurance benefits -- locally adjusted after-tax annual dollars;*

3. retirement benefits -- locally adjusted after-tax annual dollars;*

4. anticipated three-year salary -- locally adjusted after-tax annual dollars;*

5. anticipated five-year salary -- locally adjusted after-tax annual dollars;*

6. proximity to relatives -- one way jet flight time in hours;

7. degree of urbanity -- standard metropolitan area population;

8. climate -- direct worth estimate;**

9. daily commuting requirements -- one-way travel time in hours;

10. proportion of time away from home -- annual percentage;

11. duration of extended trips -- maximum trip length in days;

12. immediate training requirements -- required training time in months;

13. personal interest in the technical content of the job -- direct worth estimate;**

---

*All dollar figures were adjusted to account for differences in average living costs associated with different geographical locations in the United States.

**A direct worth point score was assigned subjectively to each alternative in this instance.

14. degree of variety implicit in the job -- direct worth estimate;* and

15. amount of training in management skills realizable from the job -- direct worth estimate;*

A pictorial display of this criterion hierarchy, complete with performance measures, is shown in Table 1. The dotted horizontal line indicates the region of demarcation between performance criteria and performance measures. The reader will notice that abbreviations are sometimes used in Table 1 to conserve space. However, review of the text should clear up any doubts about the meaning of these abbreviations.

THE CRITERION SCORES

Of the fifteen performance measures listed in The Criterion Hierarchy above and displayed in Table 1, only eleven were defined in such a manner as to require explicit scoring functions. In the remaining four instances, he decided to assign direct worth estimates to the relevant aspects of each alternative job offer. All eleven of the explicit scoring functions were sketched by a graphical technique similar to the one set forth in scoring procedure 20, Appendix T of this paper.

Table 2 below shows the estimated performance of each of the four alternatives on his fifteen performance measures.

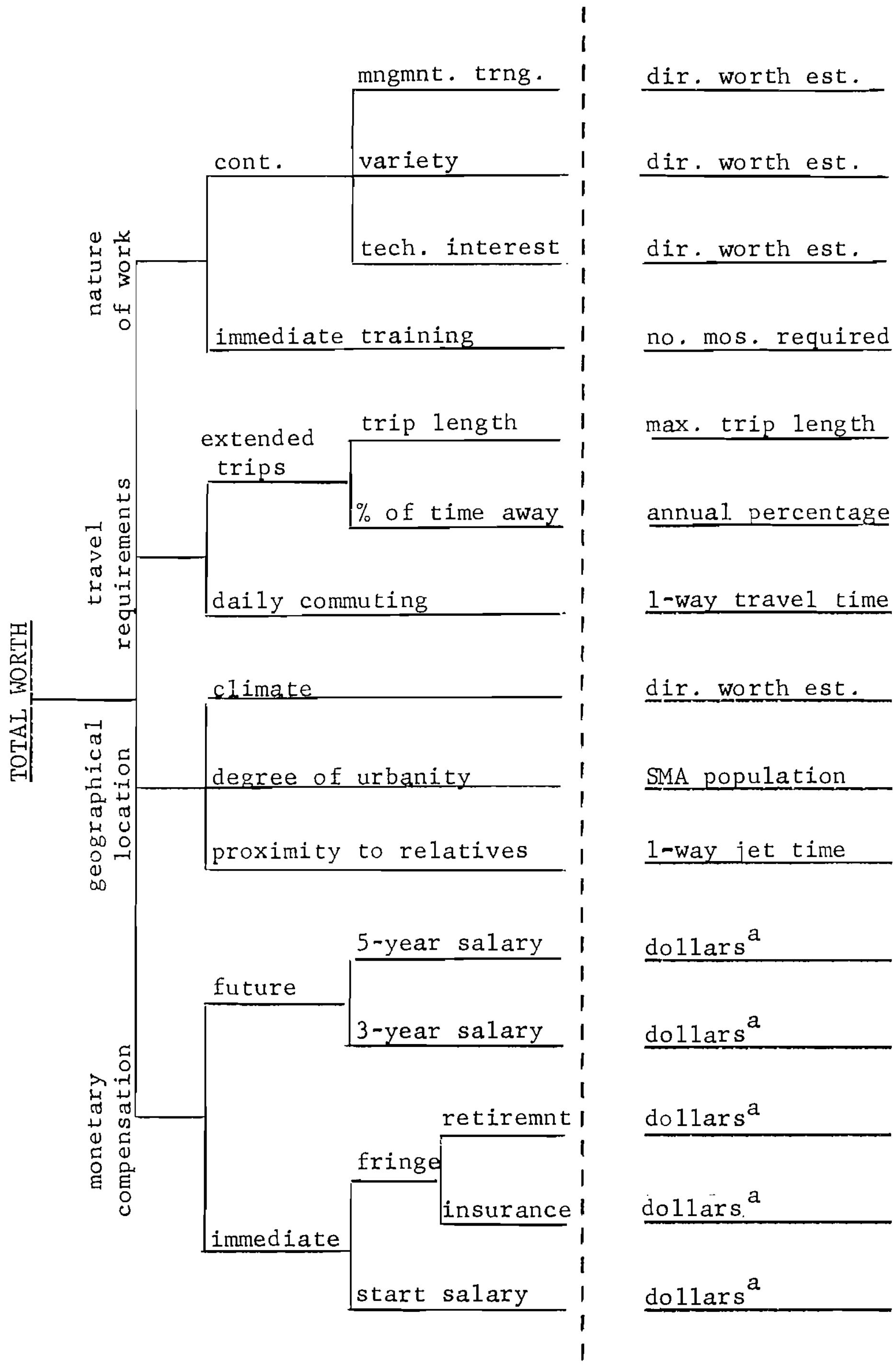
Table 3 shows the worth scores assigned either by graphical scoring functions or by direct worth estimation to the performance data associated with each alternative.

---

*A direct worth point score was assigned subjectively to each alternative in this instance.

Table 1

THE CRITERION HIERARCHY



[a]All dollar figures locally adjusted after taxes.

Table 2

ESTIMATED PERFORMANCE

Performance Criterion	Alt. I	Alt. II	Alt. III	Alt. IV
Starting salary	$ 8,100/yr.	$ 8,250/yr.	$ 8,733/64.	$8,550/yr.
Insurance benefits	$ 475/yr.	$ 550/yr.	$ 475/yr.	$ 400/yr.
Retirement benefits	$ 750/yr.	$ 1,000/yr.	$ 1,100/yr.	$ 875/yr.
Three-year salary	$11,250/yr.	$ 9,500/yr.	$10,500/yr.	$10,500/yr.
Five-year salary	$15,000/yr.	$10,500/yr.	$11,500/yr.	$11,500/yr.
Proximity to relatives	0 hrs.	0 hrs.	5 hrs.	1 hr.
Degree of urbanity	2.5 million	2.5 million	1.0 million	15.0 million
Climate	*	*	*	*
Daily commuting	.50 hrs.	1.00 hrs.	.25 hrs.	1.25 hrs.
% time away	0 %	10 %	0 %	35 %
Extended trip duration	0 days	5 days	0 days	20 days
Required job training	9.0 months	.5 months	1.0 months	.5 months
Interest in job	*	*	*	*
Variety	*	*	*	*
Training in management	*	*	*	*

*Means direct worth estimate was made.

Table 3

ASSIGNED WORTH SCORES

Performance Criterion	Alt. I	Alt. II	Alt. III	Alt. IV
Starting salary	.68	.70	.75	.73
Insurance benefits	.60	.70	.60	.50
Retirement benefits	.60	.80	.90	.70
3-year salary	.75	.63	.70	.70
5-year salary	.75	.45	.53	.53
Proximity to relatives	1.00	1.00	.10	.50
Degree of urbanity	1.00	1.00	.70	.80
Climate	.70*	.70*	.85*	.60*
Daily commuting	.60	.50	.90	.40
Percent time away	1.00	.70	1.00	.35
Extended trip duration	1.00	.70	1.00	.50
Required job training	.50	.90	.80	.90
Interest in job	.40*	.60*	.75*	.85*
Variety	.50*	.80*	.70*	.90*
Training in management	.70*	.85*	.75*	.80*

*Means direct worth estimate was made.

THE WEIGHTS

Numerical weights were then assigned to sub-criteria at every branching point in the hierarchy. For the major criteria, this process yielded the following weights:

1.	Monetary compensation	.33
2.	Geographical location	.17
3.	Travel requirements	.17
4.	Nature of work	.33
	Total	1.00

Within monetary compensation, weights were assigned as follows:

1. Immediate compensation				.70
(a) Starting salary			.90	
(b) Fringe benefits			.10	
(1) Insurance benefits		.60		
(2) Retirement benefits		.40		
Total		1.00		
Total			1.00	
2. Future compensation				.30
(a) Anticipated three-year salary			.65	
(b) Anticipated five-year salary			.35	
Total			1.00	
Total				1.00

Within geographical location, weights were assigned as follows:

1.	Proximity to relatives	.40
2.	Degree of urbanity	.40
3.	Climate	.20
	Total	1.00

Within travel requirements, weights were assigned as follows:

1. Daily commuting requirements		.20
2. Extended trips		.80
(a) Proportion of time away from home	.40	
(b) Duration of extended trips	.60	
Total	1.00	
Total		1.00

Finally, within nature of work, weights were assigned as follows:

1. Immediate training requirements		.40
2. Continuing aspects		.60
(a) Personal interest in the technical content of the job	.50	
(b) Degree of variety implicit in the job	.30	
(c) Amount of training in management skills realizable from the job	.20	
Total	1.00	
Total		1.00

The above assignment of weights lead to the following distribution of "effective" weights on each of the fifteen lowest-level performance criteria:

1.	Starting salary	.208
2.	Insurance benefits	.014
3.	Retirement benefits	.009
4.	Anticipated three-year salary	.064
5.	Anticipated five-year salary	.035
6.	Proximity to relatives	.068
7.	Degree of urbanity	.068
8.	Climate	.034
9.	Daily commuting requirements	.034
10.	Proportion of time away from home	.054
11.	Duration of extended trips	.082
12.	Immediate training requirements	.132
13.	Personal interest in the technical content of the job.	.099
14.	Degree of variety implicit in the job	.059
15.	Amount of training in management skills realizable from the job	.040
	Total	1.000

THE ADJUSTED EFFECTIVE WEIGHTS

His next step was to adjust the "effective" weights according to the perceived interpretive quality of each performance measure. This led to a set of "adjusted effective" weights which could then be applied to the worth scores shown in Table 3. The original "effective" weights, the adjusting factors, and the final set of "adjusted effective" weights are shown below in Table 4.

Table 4

"EFFECTIVE" WEIGHTS, ADJUSTING
FACTORS, AND "ADJUSTED EFFECTIVE" WEIGHTS

Performance Criterion	"Effective" Weights	Adjusting Factors	"Adjusted Effective" Weights
Starting salary	.208	1.00	.268
Insurance benefits	.014	.95	.017
Retirement benefits	.009	.95	.012
Three-year salary	.064	.75	.062
Five-year salary	.035	.75	.034
Proximity to relatives	.068	.80	.069
Degree of urbanity	.068	.75	.066
Climate	.034	.90	.040
Daily commuting	.034	.85	.037
Percent time away	.054	.50	.035
Extended trip duration	.082	.85	.090
Required job training	.132	.70	.118
Interest in job	.099	.60	.076
Variety	.059	.60	.045
Training in management	.040	.60	.031

THE TOTAL WORTH SCORES: TESTING THE ENTIRE ALGORITHM

His last step was to multiply the criterion scores by their "adjusted effective" weights and add the products to determine each alternative's total worth score. This was accompanied by tests of various subsets and the entire set of fifteen performance measures as described in Steps 7 through 11 in Section VII, Testing the Entire Assessment Algorithm. The results of these activities are shown in Table 5 below.

The testing procedure induced him to alter slightly some of his original weights and adjusting factors. However, no alterations were made due to incomplete or interdependent criteria or due to incorrect scoring functions.

Having progressed to this point in the assessment process, he had developed a slight preference for Alternative II over the other three contenders. He had been unable to discriminate among the four at the outset. In fact, it was just for this reason and the importance to him of making the right choice that he undertook formal assessment in the first place.

The figures presented in all tables reflect the end result of his testing and adjusting activities.

Inspection of Table 5 shows that Alternative II achieved the highest total worth score. As it turned out, Alternative II was selected.

Table 5

TOTAL WORTH SCORES

Performance Criterion	Alt. I	Alt. II	Alt. III	Alt. IV
Starting salary	.182	.187	.201	.195
Insurance benefits	.010	.012	.010	.009
Retirement benefits	.007	.010	.011	.008
Three-year salary	.047	.039	.043	.043
Five-year salary	.026	.015	.018	.018
Proximity to relatives	.069	.069	.007	.035
Degree of urbanity	.066	.066	.046	.053
Climate	.028	.028	.034	.024
Daily commuting	.022	.019	.033	.015
Percent time away	.035	.025	.035	.012
Extended trip duration	.090	.063	.090	.045
Required job training	.059	.106	.094	.106
Interest in job	.030	.046	.057	.065
Variety	.023	.036	.032	.041
Training in management	.022	.026	.023	.025
Total worth	.716	.747	.734	.694

IX. INTEGRATING WORTH NOTIONS WITH RESOURCE EXPENDITURES

Recall that great care was exercised throughout our assessment procedure to separate worth considerations from resource considerations. The reasons for this were explained in Section IV, subsection Legitimate Operations on Worth Points, and a procedure was presented in Section VII, subsection Identifying Resource Considerations, to implement the separation. Now we shall attempt to integrate these concepts.

Our strategy will be two-fold. First, the most salient type of resource expended will be identified, and all other types of expenditures will be re-expressed as an equivalent amount of this salient resource. Normally, monetary resources will be selected as the most salient type. If only monetary expenditures are involved, then this first step is unnecessary.

Second, a procedure for calibrating worth points in terms of equivalent monetary units (dollars) will be developed. Specifically, decision makers will be asked to indicate the most they would be willing to pay over and above current costs to obtain additional performance of various kinds over and above existing levels. "Current" costs and "existing" levels of performance will refer to whatever system alternative is currently in effect (e.g., the existing system of transportation within the Northeast Corridor).

Having obtained an equivalence between worth points and all resource expenditures, it will then be possible to compute the net worth of each alternative (i.e., equivalent worth dollars less actual cost dollars). These net worths will all be expressed as differential net worths (i.e., differential with respect to the currently existing system alternative). Differential net worth has been chosen as a summary statistic not only because it seems eminently appropriate in its own right, but also because the procedure by which worth points are converted into equivalent dollars generates differential measures. This will be demonstrated shortly.

OBTAINING DOLLAR EQUIVALENTS OF OTHER RESOURCES

The equivalent dollar cost of various other types of resource expenditures is sometimes easy to estimate by an economic analysis. Thus, equivalent dollar costs may be assigned to time expenditures in the following situations.

1. The cost of shipping delays encountered by perishable produce on its way to market may be equated with the consequent loss in revenue suffered.

2. The cost of nonproductive travel time for businessmen on their way to a conference may be estimated, on the average, as the product of hours lost times their hourly rate of compensation.

Unfortunately, however, a purely economic analysis would not be sufficient in the following situations.

1. Family income would not provide a very useful basis for estimating the substantial discomfort suffered by vacationing families delayed in transit at the height of the tourist season.

2. Children, students, housewives, and other unemployed members of the population waste a great deal of time traveling from place to place, but they lack any income to serve as a basis for estimating equivalent dollar costs.

Decision makers must once again engage in some soul-searching to establish appropriate equivalences among resource expenditures. Economic analysis can and should be used whenever possible to resolve these questions. However, as was demonstrated above, it cannot be relied upon exclusively. A suggested procedure for establishing equivalences is presented below.

Step 1. List all types of resource expenditures required to produce, operate, and use each alternative. This list will include all items identified as resources at the outset of the assessment procedure plus any items deleted from the criterion hierarchy as described in Section VII, subsection Identifying Resource Considerations.

Step 2. If all listed items are monetary expenditures, no further analysis is required. If one or more are non-monetary, take the subset of non-monetary items, and proceed to Step 3. (Note: Monetary

resources have been chosen as the most salient for purposes of this procedure. However, any other type could serve just as well.)

Step 3. Choose one of the non-monetary items on the list. Any item will do. Then identify the levels of expenditure required both on this item and on monetary cost by the current alternative.

Step 4. Choose another alternative which neither dominates nor is dominated by the current one in terms of these two resource expenditures. Clearly identifying both resource expenditures required by the current alternative, determine the maximum dollar premium which the decision maker would be willing to pay to reduce the non-monetary expenditure from its required level on the other alternative to the (assumed lower) level on the current one. Alternatively, determine the minimum cost reduction which the decision maker would require to permit increasing the non-monetary expenditure from its required level on the other alternative to the (higher) level on the current one.

Step 5. Repeat Step 4 until all alternatives have been exhausted.

Step 6. Repeat Steps 3 through 5 until all non-monetary resources have been exhausted.

Step 7. The above steps should generate a locus of indifference points relating changes in each of the non-monetary resources to changes in dollar cost. If enough points are available, a smooth curve may be drawn through them to determine an equivalence function. The method of least-squares may be invoked to obtain a more precise fit, if desired. More decision makers may be asked to go through the above procedure to obtain larger sample sizes for each locus of points.

This completes the procedure.

OBTAINING DOLLAR EQUIVALENTS OF WORTH POINTS

Dollar equivalents for worth points will be sought in a similar, but slightly more complicated manner. It is assumed that a complete assessment has been made of all feasible alternatives and that worth

scores of the type displayed in Section VIII, Table 5, have been computed. The following steps are designed to convert these scores into equivalent dollar values.

Step 1. Choose a manageable number of (e.g., five to ten) lowest-level performance criteria satisfying all of the following requirements:

a. "Adjusted effective" weights are at least moderately high (i.e., fall at or above the median of all "adjusted effective" weights).
b. Explicit scoring functions have been defined (i.e., ignore criteria for which direct worth estimates must be made).
c. Estimates have been made on all corresponding performance measures for all feasible alternatives.
d. Scoring functions are not excessively flat throughout the performance range determined by the feasible alternatives.

Step 2. Identify the subset of feasible alternatives which neither dominate completely nor are completely dominated by the current alternative in terms of whichever performance measures were singled out in Step 1. Hopefully, this subset will include all feasible alternatives except the current one. If it does not, Steps 1 and 2 may be repeated until the largest possible subset of alternatives has been achieved.

Step 3. Choose one of the performance measures singled out in Step 1. Any one will do. Then identify the level of performance attained on this measure by the current alternative. Identify also the relevant cost associated with the current alternative.*

Step 4. Choose one of the other alternatives set aside in Step 2. Any one will do. Clearly identifying both cost and measured

---

*Relevant cost here means relevant to obtaining the designated type of performance. Thus, if some aspect of passenger comfort were being measured, passenger fare would be the relevant cost. If some aspect of reliability or convenience in freight shipments were being measured, shipping charges would be the relevant cost.

performance associated with the current alternative, determine the maximum dollar premium which the decision maker would be willing to pay to improve performance from its level on the current alternative to the (assumed preferred) level on the other one. Alternatively, determine the minimum cost reduction which the decision maker would require before accepting a deterioration of performance from its level on the current alternative to the (inferior) level on the other one.

Step 5. Repeat Step 4 until all alternatives have been exhausted.

Step 6. Repeat Steps 3 through 5 until all performance measures have been exhausted.

Step 7. Convert each of the performance changes generated in Steps 3 through 6 into equivalent changes in total worth scores. This can be accomplished by reading each point score difference off the appropriate scoring function and multiplying this difference by the corresponding "adjusted effective" weight.

Step 8. The above steps should generate a locus of indifference points relating changes in total worth to changes in dollar cost. Enough points should be available so that a smooth curve may be drawn through them. This will define an equivalence function. The method of least-squares may be invoked to obtain a more precise fit, if desired. More decision makers may be asked to go through the above procedure to obtain a larger number of points and to check for consensus.

This completes the procedure.

CONCLUDING REMARKS

The reader may wonder at this point why the foregoing pair of procedures was not adopted at the outset of formal assessment. Why, in fact, was it necessary to wade through the arduous and time-consuming tasks of criterion structuring, scoring, weighting, and testing, when all point scores are eventually converted into dollars? Why

not assign dollar values in the first place? The answer is quite straightforward.

Past research (by the writer) has shown that decision makers find it exceedingly difficult, if not impossible, to make meaningful trade-offs in dollar terms (or any other terms, for that matter) without having first made a substantial effort to create a consistent worth structure. Our complete assessment procedure serves to induce and to guide just such an effort. If dollar trade-offs are requested at the outset, decision makers tend to answer only those questions for which a clear economic justification can be found. But this is much too restrictive. A "complex" decision problem, like choosing transportation systems, should not be based solely on factors for which economics provides a clear answer. Existing economic models are just plain inadequate to calibrate many of the crucial "intangibles" in dollar terms. By making it clear at the outset that decision makers must think through all relevant factors, that they must create a quantitative unit of worth for all of these factors out of their own imagination and experience, and that this unit of worth need not (initially) bear any direct relation to dollars, a far more comprehensive assessment should result. This is the premise on which our procedure rests.

On the other hand, total worth scores are converted into equivalent dollars at the end for three reasons. First, this permits computing a net worth in dollars for each alternative. Without such a conversion, no single measure of net worth could be achieved. Second, conversion proceeds from points to dollars (rather than from dollars to points) because dollars are easily understood by the world at large, while worth points will only have meaning to those decision makers actually involved in assessment. This makes it far easier to communicate one's final conclusions regarding the decision at hand. Third, having net worth statistics calibrated in terms of differences from the current alternative will permit a very high-order summarization of conclusions in both a static and a dynamic

context. These topics will be treated in the next section.

One final comment. Since the two procedures presented in Section IX, subsections Obtaining Dollar Equivalents of Other Resources and Obtaining Dollar Equivalents of Worth Points, generate trade-offs expressed as differences from the current alternative, and since these differences are localized by the ranges of performance and resource expenditures determined by the set of feasible alternatives, it is reasonable to expect that all trade-off functions may be linearly approximated. Furthermore, it is quite reasonable to restrict these functions so that decision makers would pay nothing extra for no extra worth. This suggests that all smooth curves, either drawn by hand or fitted by least-squares, may be straight lines passing through the origin of the "differential cost"-"differential worth" axes.

X. TOWARD AN ULTIMATE DECISION RULE

The entire discussion to date has assumed that whatever alternatives are proposed have been designed to perform a single, well-defined job. In the case of the complete example presented in Section VIII, the job was "to obtain initial employment for a student upon graduation from a business school." In the case of the partial example referenced in Section VIII, the job was "to transport an average, middle-income businessman from Washington to New York." We shall now attempt to move toward an ultimate decision rule for choosing among proposed alternatives. To accomplish this, however, we must first consider two additional issues omitted from previous discussion. Specifically:

1. Not all individuals and interest groups affected by whichever alternative is eventually chosen share the same definition of what job that alternative is supposed to do for them; and

2. Job definitions, whether or not shared at a given point in time, are very likely to change as technology, tastes, habits, and aspiration levels change.

Our ultimate decision rule must be sufficiently comprehensive to reflect both between-group differences (a political problem) and temporal changes (a psychological and behavioral problem).

REFLECTING BETWEEN-GROUP DIFFERENCES

Clearly, the job to be performed by a transportation system is perceived quite differently by:

1. The profit-motivated operator and his fare-paying passengers;

2. The zealous chamber of commerce president and the disgruntled homeowner displaced by eminent domain;

3. The strawberry producer seeking wider markets for his perishable product and the inefficient monopolist whose livelihood depended upon sole access to a previously inaccessible region; and

4. The vacationing family, complete with screaming babies and howling pets, and the elder statesman attempting to organize his thoughts prior to an important diplomatic confrontation.

Besides reflecting just different perceptions, the above examples were intentionally selected to highlight the many and real opportunities for conflicting perceptions.

Now any decision rule should at the very least lay bare for careful scrutiny substantial between-group differences. Resolving such differences is a separate and much more difficult task. We shall undertake only the former task in this section. A way of assessing and displaying different perceptions of the same alternative in a form convenient for between-group trade-off analyses is suggested below.

Step 1. Partition the entire population of individuals affected by the decision into relatively homogeneous interest groups. Homogeneous, here, means likely to share similar job definitions and similar worth judgments with respect to the decision at hand.

Step 2. Identify and set aside those groups, if any, whose interests are to be minimally satisfied (see discussion in Section VII, subsection Establishing Major Objectives).

Step 3. Prepare a summary net worth table as follows.

a. Lay out all feasible alternatives along the rows of the table, starting with the current alternative in the first row.
b. Lay out the various groups whose interests are to be optimized (i.e., all groups except the ones set aside in Step 2) along the columns. Although not critical, it will be helpful to sequence these groups from left to right in descending order of perceived importance. One interest group might be perceived as more important than another because it is larger, it is politically more influential, it is more deserving of governmental support, etc. By so sequencing interest groups, primary attention may later be focused upon the left-most columns of the table.
c. Enter zeros along the entire first row of the table. Since all subsequent entries will be net worths computed as differences from the current alternative, and because the

current alternative never differs from itself, zeros are required here.

d. Using the indifference functions plotted in Section IX, subsection Obtaining Dollar Equivalents of Worth Points, compute the differential net worth of each alternative for each interest group, and enter the results in the appropriate cells of the table.*

e. These differential net worths are normally computed for an individual member of each interest group. This was the strategy followed in both examples discussed in Section VIII. On the assumption that gratifying any one member of a homogeneous group is just as important as gratifying any other member, all entries in each column should be multiplied by the number of individuals contained in the corresponding interest group. Such a simple technique would probably not be appropriate across heterogeneous and frequently conflicting interest groups.

Step 4. Delete any alternatives (rows) which are completely dominated by any other alternative(s). That is, inspect every possible pair of rows (there will be $\frac{1}{2} N^2 - \frac{1}{2} N$ pair-wise comparisons for N rows), and look for any row whose entries are uniformly smaller than the corresponding entries of the other. Although it is unlikely that any such rows will be found, unless there are only a few columns in the table, it is worthwhile making this check. Such rows (alternatives), if located, may be eliminated straightaway from further consideration. It would never be sensible to select an alternative regarded as inferior from everybody's point of view.**

---

*Strictly speaking, computation of each column of entries requires a separate application of the complete assessment procedure (once for each interest group). However, this is not as difficult as it may seem. Large segments of any worth structure formulated for one interest group would apply equally to many other groups.

**The RAND group responsible for generating the partial example referenced in Section VIII initially partitioned the population of passengers into about twenty separate interest groups (this number was later reduced); and this did not consider freight, operator, or societal groups. Consequently, the likelihood of detecting any dominated alternatives would appear slight.

Step 5. Re-check the remaining alternatives for feasibility. That is, insure that they all satisfy every mandatory performance requirement, including minimal satisfaction of the interest groups set aside in Step 2. Delete infeasible alternatives (rows), if any are found.

Step 6. The resulting table summarizes differential net worths for all feasible and non-dominated alternatives as separately perceived by every whole interest group.

This completes the procedure.

REFLECTING TEMPORAL CHANGES

Reflecting temporal changes means constructing separate summary net worth tables for successive points in time. The first such table would reflect current conditions at the time of assessment. Subsequent tables would reflect conditions assumed to hold at various future times. But this raises two questions.

1. How can future table entries be estimated?
2. How would this additional information be useful, assuming it could be generated?

Regarding the first question, procedures have already been developed to provide much of the data required for future net worth tables. These procedures seek to predict temporal shifts in aspiration levels resulting from improved technology. Since this work is still in a preliminary stage, it will not be reported here. We shall proceed instead to the second question and discuss how time-series information could be utilized for decision making purposes.

Assuming we have successive net worth tables, a reasonable procedure would be to cumulate corresponding entries over time. We might also choose discount rates to reflect both preferences for receiving positive net worths sooner rather than later and reduced aversions to negative net worths, if they occur far in the future. The cumula-

tive sum of discounted net worths could then be interpreted as a present worth or present value statistic. Whether or not future net worths are discounted, a plot of this cumulative statistic against time would serve to illustrate the sensitivity of final results to whatever planning horizon (i.e., number of time periods) is being considered. It would also permit another and more meaningful form of dominance analysis. A procedure for obtaining these results is presented below.

Step 1. Excluding the first row of entries, which will contain zeros everywhere in every table, choose some entry. Any one will do.

Step 2. If discounting is deemed inappropriate, proceed immediately to Step 9. If discounting is deemed appropriate, identify the interest group (column) associated with the chosen entry.

Step 3. Choose a discount rate to reflect how intensely that group desires immediate versus postponed gratification.*

Step 4. Compute as many discount factors for that group as there are successive net worth tables. The formula for successive discount factors is

$$DF_{gt} = \frac{1}{(1 + r_g)^t} ,$$

where DF_{gt} is the discount factor for group g in period t, r_g is the the annual discount rate assigned to group g, and t is a serial integer ranging from 0 (in the current year) to N - 1, assuming N tables spaced at annual intervals.

*A separate discount rate is defined for each interest group to reflect the real between-group differences in contemporary society. Thus, a militant minority group might crave immediate gratification, implying a high discount rate, while an established business group might be satisfied with a lower discount rate equal to their average rate of return on investment.

Step 5. Replace the chosen entry in each successive table with the product of itself times the corresponding discount factor computed in Step 4.

Step 6. Repeat Step 5 until all entries in the column chosen in Step 2 have been exhausted.

Step 7. Choose an entry from a different column, and return to Step 2. Repeat Steps 2 through 7 until all groups (columns) have been exhausted.

Step 8. Return to the entry chosen in Step 1.

Step 9. Replace the chosen entry in each successive table with the sum of itself and the corresponding entry in the immediately preceding table. In the case of the first table, each entry is left unchanged (i.e., replaced with itself).

Step 10. Repeat Step 9 until all entries, except those along the first row, have been exhausted. The result is a sequence of tables, each containing cumulative differential (possibly discounted) net worth entries.

Step 11. Choose an interest group (column). Any one will do.

Step 12. Choose an alternative (row). Once again, any one will do. These two choices together define a specific table entry.

Step 13. Plot on graph paper the locus of corresponding entries in successive tables (vertical axis) against the serial integer assigned to each table (horizontal axis).

Step 14. Connect the sequence of plotted points with straight lines.

Step 15. Repeat Steps 12 through 14 until all alternatives (rows) have been exhausted. Use the same piece of graph paper for all plots associated with the interest group chosen in Step 11. This will generate a dynamic and pictorial representation of cumulative differential net worths associated with every alternative as viewed

by the chosen interest group.*

Step 16. Find a fresh piece of graph paper, and repeat Steps 11 through 15 until all interest groups (columns) have been exhausted. This will generate a separate display for each interest group.

Step 17. It is now a simple matter to identify dominated alternatives, at least for a given interest group, by visual inspection of the graphs just generated. Any alternative whose plotted points lie entirely below the plotted points of at least one other alternative is completely dominated. It is, therefore, of no interest to that particular group.

This completes the procedure.

MAKING THE FINAL DECISION

There remains one last problem to solve before a complete decision rule can be formulated. Once formulated, the final decision can be made. Unfortunately, however, this last problem constitutes the most difficult aspect of the entire assessment process. How do we trade-off among different and frequently conflicting interest groups?

One simple, but probably inappropriate, way to answer this question would be to treat all interest groups as equally important. Then, the rows of every table created in Section X, subsection Reflecting Temporal Changes, could be added to obtain a single score for each alternative in every period. These results could be transferred to a single sheet of graph paper as previously described. The final decision would then depend only on the time horizon. For any given time horizon, whichever alternative had the highest cumulative differential (possibly discounted) net worth score would be judged superior. If any alternative possessed a higher score in all time horizons, it could be chosen straightaway.

---

*The current alternative will appear on this graph as a sequence of points along the horizontal axis. This will be true no matter which interest group was selected, since all first-row entries are zero.

A slight modification of the above procedure would improve its reasonableness. Decision makers might assign weights to the various interest groups (see weighting procedure in Section VII, subsection Assigning Weights). These weights would indicate the relative importance of satisfying each group's interests vis-a-vis the others. Then, a weighted sum of row entries could be computed in place of the simple sum described above. The final decision would depend jointly upon these weighted sums and the planning horizon chosen, just as before.

The reader can undoubtedly think of other analytical devices for reducing whole rows of table entries into a single number. But this is really not the point. We are dealing here with a problem which is essentially political -- not analytical. Coming up with a solution which is seen as equitable, politically defensible, practically workable, and at least minimally acceptable to all interest groups is not the job of the analyst. The moment of truth has come for the decision maker. This last decision he must make for himself.

XI. A CRITICAL REVIEW

In this section we shall briefly review the assessment procedure. Critical scrutiny will be directed toward the methodology in an attempt to pinpoint "soft spots." The task of assigning weights, one of the more difficult parts of the procedure, will be discussed in some detail.

The reader may have noticed that several important aspects of the overall task of assessment were either ignored completely or else given only a cursory treatment. Methodological issues falling into this category include:

1. the problem of describing adequately and accurately the job to be performed by whichever alternative is finally selected -- this was ignored completely;
2. the problem of producing alternatives to accomplish the stated job -- this was also completely ignored;
3. the problem of predicting both performance and resource consequences associated with each produced alternative -- very little was said about this issue;
4. the problem of validating the descriptive accuracy of performance and resource estimates -- this was ignored completely;
5. the problem of establishing feasibility constraints (i.e., mandatory performance and/or resource requirements) on alternatives -- this issue was also ignored;
6. the problem of incorporating risk/uncertainty considerations -- this was discussed only briefly; and
7. the problem of selecting appropriate personnel to assess alternatives and to make a final choice -- except to point out that final results could depend critically upon both the identity of decision makers and the point in time when an assessment is made, this issue was largely ignored.

Now it is not claimed that the above issues are unimportant. Quite to the contrary, they are all very important, and they deserve the same amount of attention accorded to worth assessment. However,

the scope of this paper was not intended to cover these issues, except insofar as they provided a context in which to discuss worth assessment.

One critical assumption about the manner in which decision makers can be induced to formulate worth structures deserves special attention. This involves the weights. Can decision makers be comfortable with a linear weighting scheme? A typical first reaction to this question is negative on the grounds that strict linearity is too simple and restrictive. However, after realizing that linearly weighted performance criteria do not necessarily imply an assessment algorithm linear in performance measures (recall that scoring functions can assume any desirable non-linear shape), decision makers will generally retract their objection. At least this is what occurred during the experiment.

A more serious problem with the weights lies in their abstractness. In both the experiment and the partial transportation example decision makers complained that no firm basis existed for assigning weight numbers. Unlike scoring functions, where decision makers could think concretely about physical performance, weights had to be assigned on a purely intuitive basis. There are several ways to check assigned weights for reasonableness and consistency, and these will be discussed shortly. However, none of these will transform the weights into concrete entities. They will only serve to increase confidence in whatever numerical values have been assigned.

The first and most basic check on assigned weights is already built into our procedure. The testing process described in Section VII, subsection Testing the Entire Assessment Algorithm, can be and frequently is used as means of "initial tuning." The weights can be altered selectively until the rank-order of total worth scores is brought into alignment with a decision maker's subjective ordering of the alternatives. If "fine tuning" is desired in addition, an indifference analysis like the ones suggested in Section X might be performed. This would increase confidence in the cardinal significance of the weights, as well as in their ordinal significance. Still further checks on their cardinal significance could be obtained

via constrained linear regression. That is, decision makers could assign total worth scores subjectively to each alternative, and these scores could be regressed on the outputs of all scoring functions. If the regression coefficients were constrained to be non-negative with unit sum, the computed regression coefficients could then be compared directly with the weights assigned, respectively, to each scoring function. Naturally, none of these refinements need be carried out unless the final decision is highly sensitive to the weights themselves. Both the indifference analysis and the regression analysis suggested above should be preceded by a sensitivity analysis centered upon whichever weights emerged from Section VII, subsection Testing the Entire Assessment Algorithm.

One last point deserves clarification before closing. Our procedure has assumed throughout that a fixed set of discrete alternatives has been produced and that the only problem is to choose one. All testing and tuning operations, all indifference analyses, and all steps described in Section X require validated performance data. But what happens if no alternatives yet exist? What if our problem is to design alternatives rather than to assess them? The procedure is still applicable in principle (excluding the specific operations just mentioned), but the confidence we can muster in its results is substantially reduced. The experiment reported in Appendix U pointed up very clearly the importance to decision makers of having concrete alternatives before them as a check against their worth judgments. Without such checks, the experimental subjects gained very little from the procedure and lacked conviction in their assumptions, their judgments, and their final choices. Hence, using the procedure to guide design decisions could be dangerous unless carried out with extreme care.

Appendix A

SCORING PROCEDURE 1

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. two-level scale;
3. level 1 = absence of some desirable performance attribute; and
4. level 2 = presence of that desirable attribute.

Step 1. Assign zero worth points to absence of the desirable attribute.

Step 2. Assign one worth point to presence of that desirable attribute.

This completes the procedure.

Appendix B

SCORING PROCEDURE 2

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. two-level scale;
3. level 1 = absence of some desirable performance attribute; and
4. level 2 = presence of that desirable attribute in conjunction with some qualitative measure of relative worth, when present.

Step 1. Assign zero worth points to absence of the desirable attribute.

Step 2. Identify all feasible alternatives which promise the desirable attribute.

Step 3. Assemble one or more decision makers.

Step 4. After discussing collectively the various merits of the desirable attribute -- why it is important and what benefits its presence provides -- have each decision maker make a separate and independent judgment of the extent to which each feasible alternative's promised attribute satisfies the related lowest-level performance criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each feasible alternative.

Step 5. To determine each feasible alternative's score on this performance measure, assign either zero points (if the attribute is absent) or the arithmetic mean (possibly weighted) of the individual scores assigned judgmentally by separate decision makers.

This completes the procedure.

Appendix C

SCORING PROCEDURE 3

Reference in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. three, four, or five levels on the scale; and
3. all levels constitute strictly nominal categories.

Step 1. Since the scale of the physical performance measure is strictly nominal, a preference or worth ordering must be placed directly on the nominal categories. This will be accomplished by means of the same ranking procedure used in defining weights and presented in Section VII, subsection Assigning Weights.

Step 2. Assemble one or more decision makers.

Step 3. After discussing collectively the various merits of nominal categories, have each decision maker perform a separate and independent rank-ordering of the various categories. This may be accomplished by performing Steps 4 through 7 below.

Step 4. List the nominal categories in approximate order of decreasing worth, starting with the category perceived as most valuable at the top of the list. The category perceived as least valuable should appear at the bottom of the list. It is not necessary to have the categories perfectly ranked or ordered on this first pass, since subsequent operations will be performed to guarantee complete ordering.

Step 5. Compare the first two categories on the list.

a. If the first category is perceived as more valuable than the second, proceed directly to Step 6.
b. If both categories are perceived as roughly equal in worth, proceed directly to Step 6.

c. If the second category is perceived as more valuable than the first, invert their positions on the list (i.e., place the first category where the second used to be on the list, and vice versa), and then proceed to Step 6.

Step 6. Compare the lower-ranked category from Step 5 with the next-lower category on the list. Repeat the comparisons and stipulated operations in Step 5 on this new pair of categories. Continue in this manner all the way down to the end of the list until pairwise comparisons have been made between all contiguous criteria.

Step 7. After the list has been completely exhausted, go back and determine whether any inversions (position changes) occurred. If none occurred, proceed directly to Step 8. If one or more occurred, return to the head of the list, and repeat the entire procedure described in Steps 5 and 6.

Step 8. Next, have each decision maker make a separate and independent judgment of the extent to which each ranked category satisfies the related lowest-level criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each scale category.

Step 9. Finally, to determine each nominal category's point score, compute and record the (possibly weighted) arithmetic mean of the individual category scores assigned by separate decision makers in Step 8.

This completes the procedure.

Appendix D

SCORING PROCEDURE 4

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. three, four, or five levels on the scale; and
3. the scale is ordered.

Step 1. List the ordered levels in a single column.

Step 2. Inspect the level appearing at the head of the column. Is that the most preferred or the least preferred level? If most preferred, proceed to Step 4. If least preferred, proceed to Step 3.

Step 3. Invert the column, and list the levels again -- this time in reverse order. Proceed to Step 4.

Step 4. The discrete levels should now be listed in perfect order of descending relative worth. Inspect to verify that this is true. If so, proceed to Step 5. If not, check earlier steps to insure that no errors occurred. If no errors occurred, this particular performance measure should be treated by scoring procedure 20 in Appendix T.

Step 5. Assemble one or more decision makers.

Step 6. After discussing collectively the various merits of nominal categories and (hopefully) agreeing on their rank-order, have each decision maker record a separate and independent judgment of the extent to which each nominal category satisfies the related lowest-level performance criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each scale category.

Step 7. To determine each nominal category's score, compute and record the (possibly weighted) arithmetic mean of the individual category scores assigned by separate decision makers.

This completes the procedure.

Appendix E

SCORING PROCEDURE 5

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. more than five levels on the scale; and
3. all levels constitute strictly nominal categories.

Step 1. Since the scale of the physical performance measure is strictly nominal, a preference or worth ordering must be placed directly on the nominal categories. This will be accomplished by means of the same ranking procedure used in defining weights and presented in Section VII, subsection Assigning Weights.

Step 2. Assemble one or more decision makers.

Step 3. After discussing collectively the various merits of nominal categories, have each decision maker perform a separate and independent rank-ordering of the various categories. This may be accomplished by performing Steps 4 through 7 below.

Step 4. List the nominal categories in approximate order of decreasing worth, starting with the category perceived as most valuable at the top of the list. The category perceived as least valuable should appear at the bottom of the list. It is not necessary to have the categories perfectly ranked or ordered on this first pass, since subsequent operations will be performed to guarantee complete ordering.

Step 5. Compare the first two categories on the list.

a. If the first category is perceived as more valuable than the second, proceed directly to Step 6.
b. If both categories are perceived as roughly equal in worth, proceed directly to Step 6.

c. If the second category is perceived as more valuable than the first, invert their positions on the list (i.e., place the first category where the second used to be on the list, and vice versa), and then proceed to Step 6.

Step 6. Compare the lower-ranked category from Step 5 with the next-lower category on the list. Repeat the comparisons and stipulated operations in Step 5 on this new pair of categories. Continue in this manner all the way down to the end of the list until pairwise comparisons have been made between all contiguous criteria.

Step 7. After the list has been completely exhausted, go back and determine whether any inversions (position changes) occurred. If none occurred, proceed directly to Step 8. If one or more occurred, return to the head of the list, and repeat the entire procedure described in Steps 5 and 6.

Step 8. Inspect adjacent pairs of ranked scale categories. Locate that adjacent pair of scale categories which seem closest to one another in terms of their perceived worth (i.e., locate the most equally valuable adjacent pair of scale categories). Collapse these two categories into a single category.

Step 9. Repeat Step 8 as many times as is required to reduce the number of resulting categories to five. Then proceed to Step 10.

Step 10. Next, have each decision maker record a separate and independent judgment of the extent to which each ranked category satisfies the related lowest-level criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each scale category.

Step 11. Finally, to determine each nominal category's point score, compute and record the (possibly weighted) arithmetic mean of the individual category scores assigned by separate decision makers in Step 10.

This completes the procedure.

Appendix F

SCORING PROCEDURE 6

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound; and
7. constant rate of change of worth with increases in performance.

The above seven characteristics describe completely a linear scoring function passing through the origin and with positive slope equal to the reciprocal of the logical upper bound. The equation of this scoring function is

$$\text{worth score} = \frac{\text{Measured Performance}}{\text{Logical Upper Bound}}$$

A graphical picture of this scoring function appears below in Fig. 1.

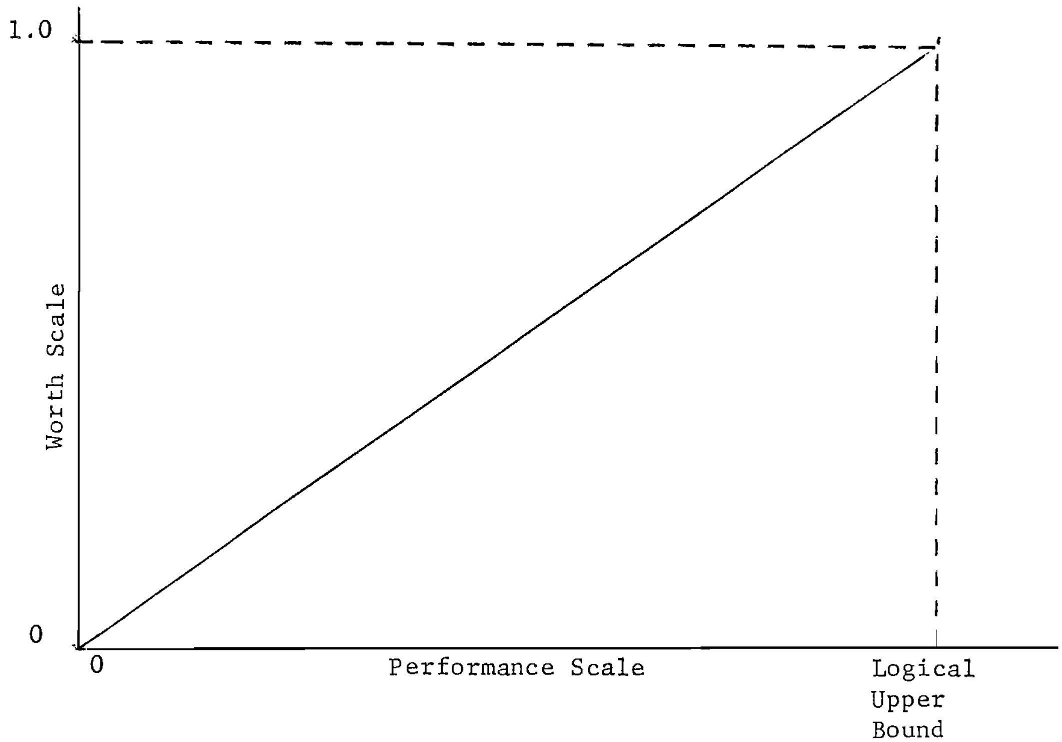


Figure 1

This completes the procedure.

Appendix G

SCORING PROCEDURE 7

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound; and
7. uniformly accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 2.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with positive second derivative (indicating uniform acceleration).

2. The minimum of the quadratic function falls exactly at the origin.

3. The upper tail of the scoring function passes through the point whose coordinates are (performance = logical upper bound, worth score = one).

These three assumptions completely determine a scoring function (see Fig. 2) whose equation is

$$\text{worth score} = \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2 .$$

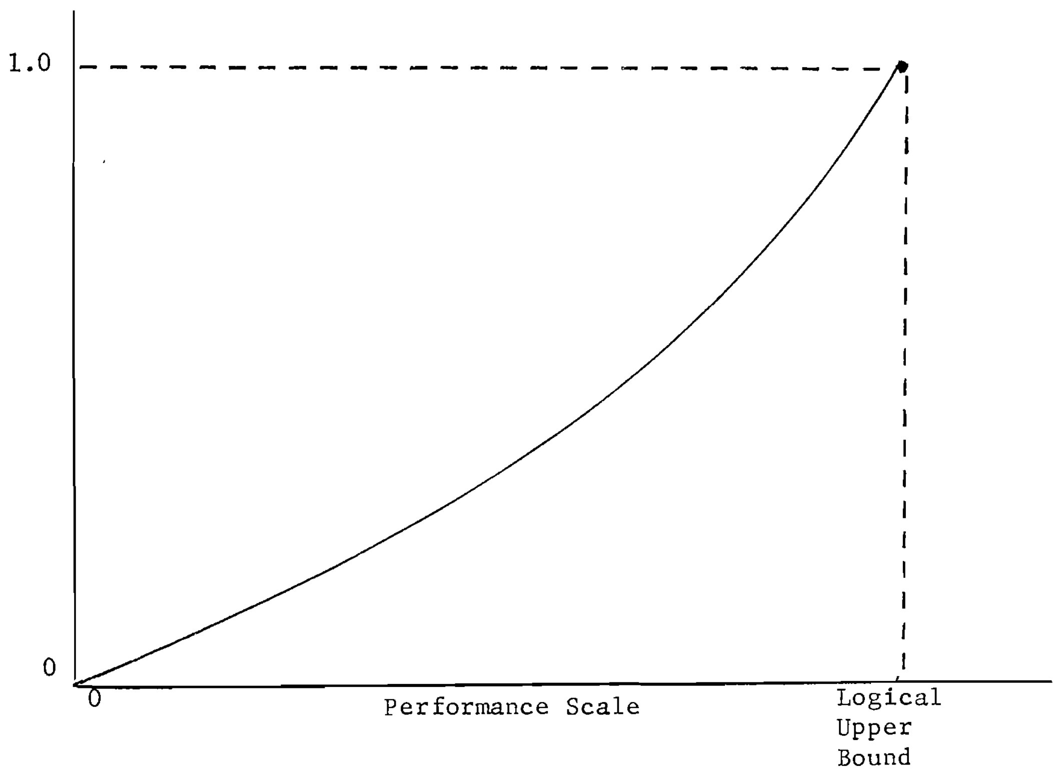


Figure 2

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix H

SCORING PROCEDURE 8

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound; and
7. uniformly decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 3.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with negative second derivative (indicating uniform deceleration).

2. The maximum of the quadratic function falls exactly at the point whose coordinates are (performance = logical upper bound, worth score = one).

3. The quadratic function passes through the origin.

These three assumptions completely determine a scoring function (see Fig. 3) whose equation is

$$\text{worth score} = 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) - \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2.$$

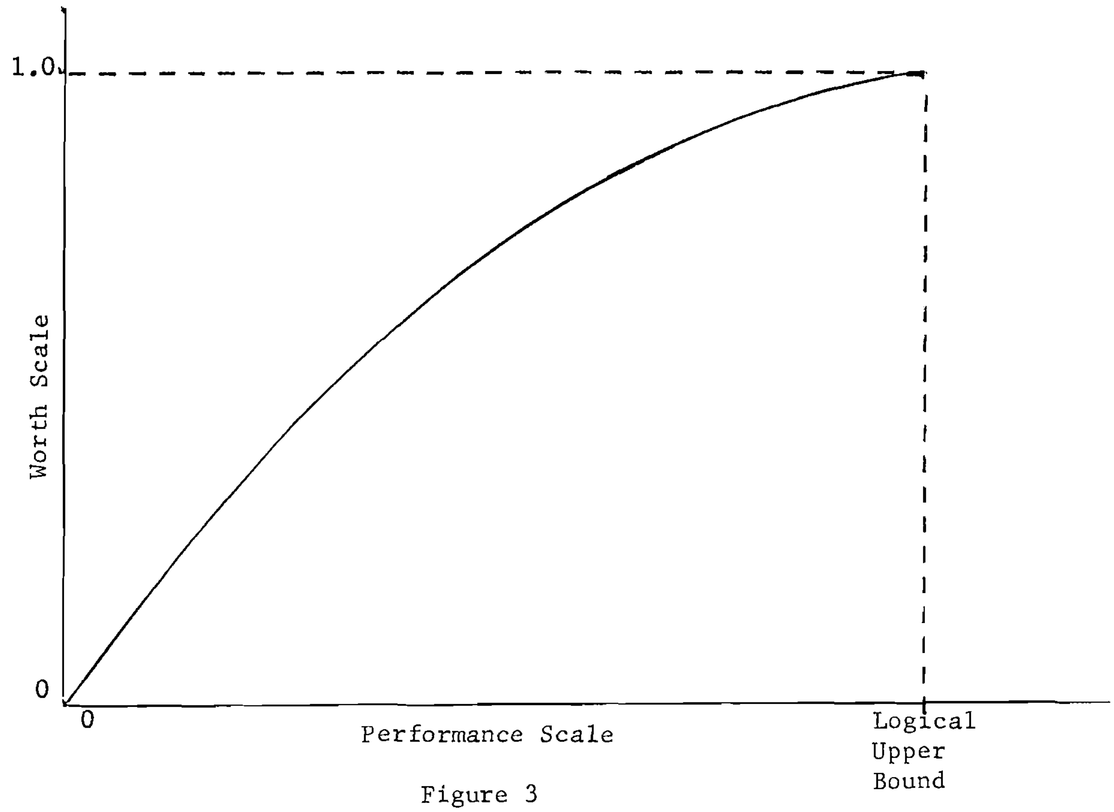


Figure 3

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix I

SCORING PROCEDURE 9

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound; and
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 4.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = \tfrac{1}{2} - \tfrac{1}{2}\,\text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right]$$

where $\pi = 3.1416$,
and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

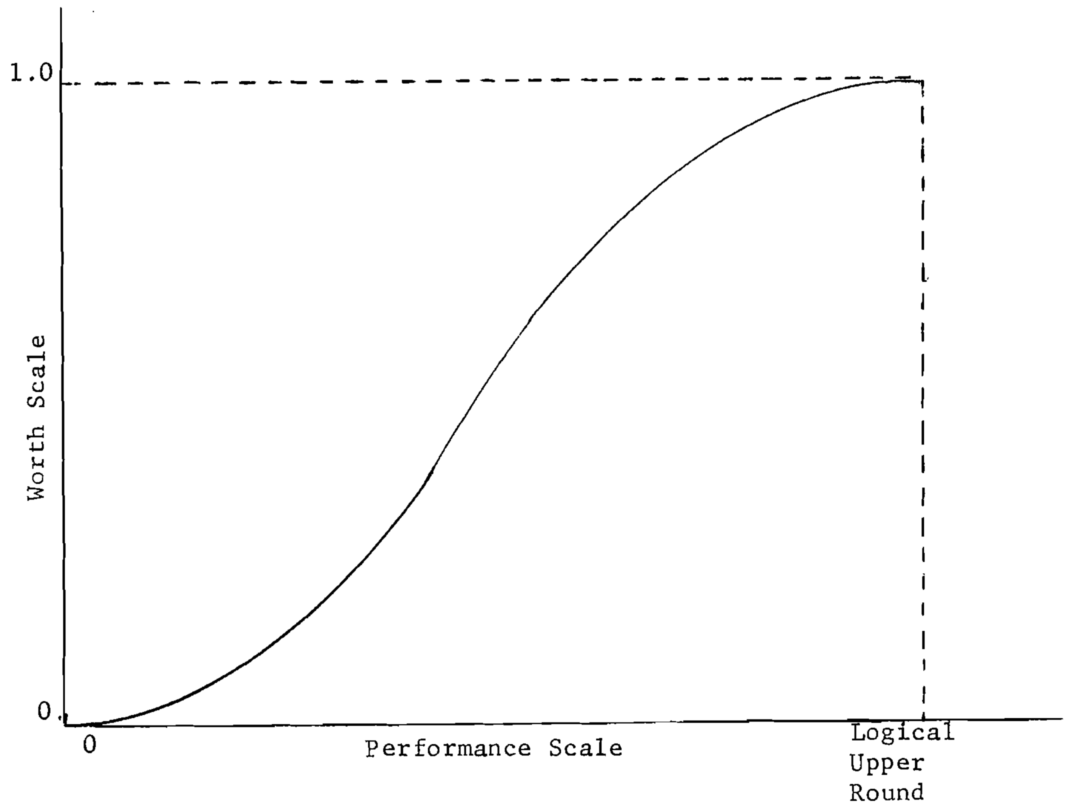


Figure 4

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix J

SCORING PROCEDURE 10

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound; and
7. first decelerating, then accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 5.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) + \tfrac{1}{2}\ \text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right] -$$

where $\pi = 3.1416$,

and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

/3

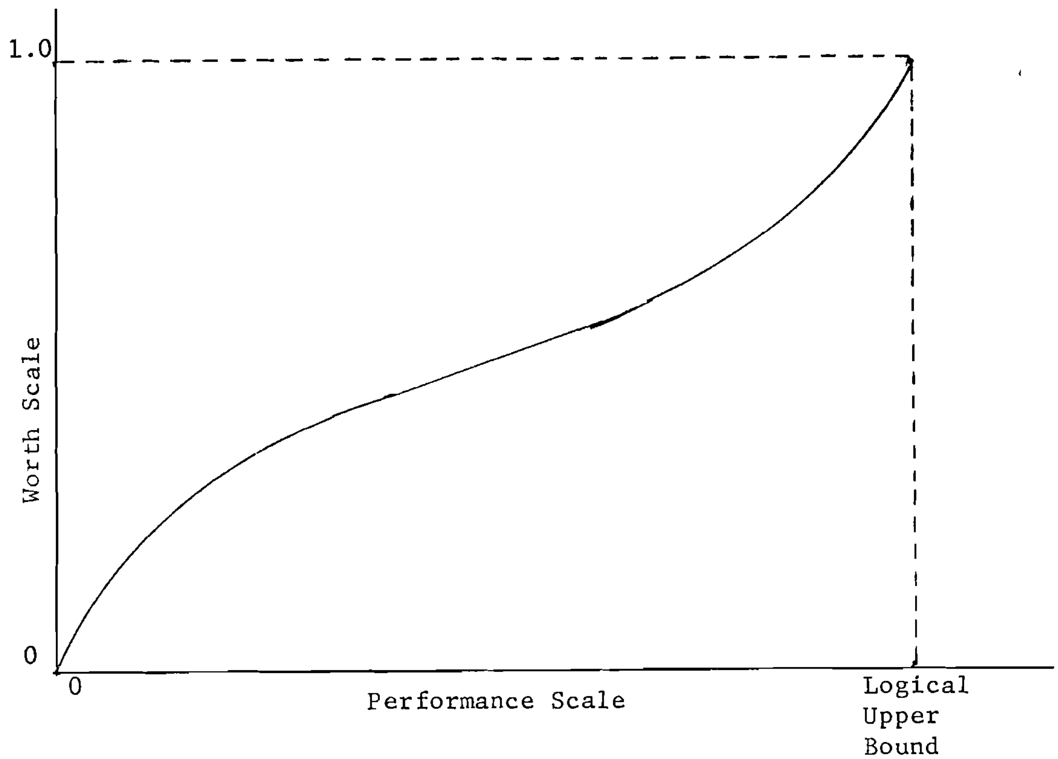


Figure 5

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix K

SCORING PROCEDURE 11

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance; and
7. constant rate of change of worth with increases in performance.

The above seven characteristics describe completely a linear scoring function passing through the point whose coordinates are (performance = zero, worth score = one) and with negative slope equal to minus the reciprocal of the logical upper bound. The equation of this scoring function is

$$\text{worth score} = 1 - \frac{\text{Measured Performance}}{\text{Logical Upper Bound}}.$$

A graphical picture of this scoring function appears below in Fig. 6.

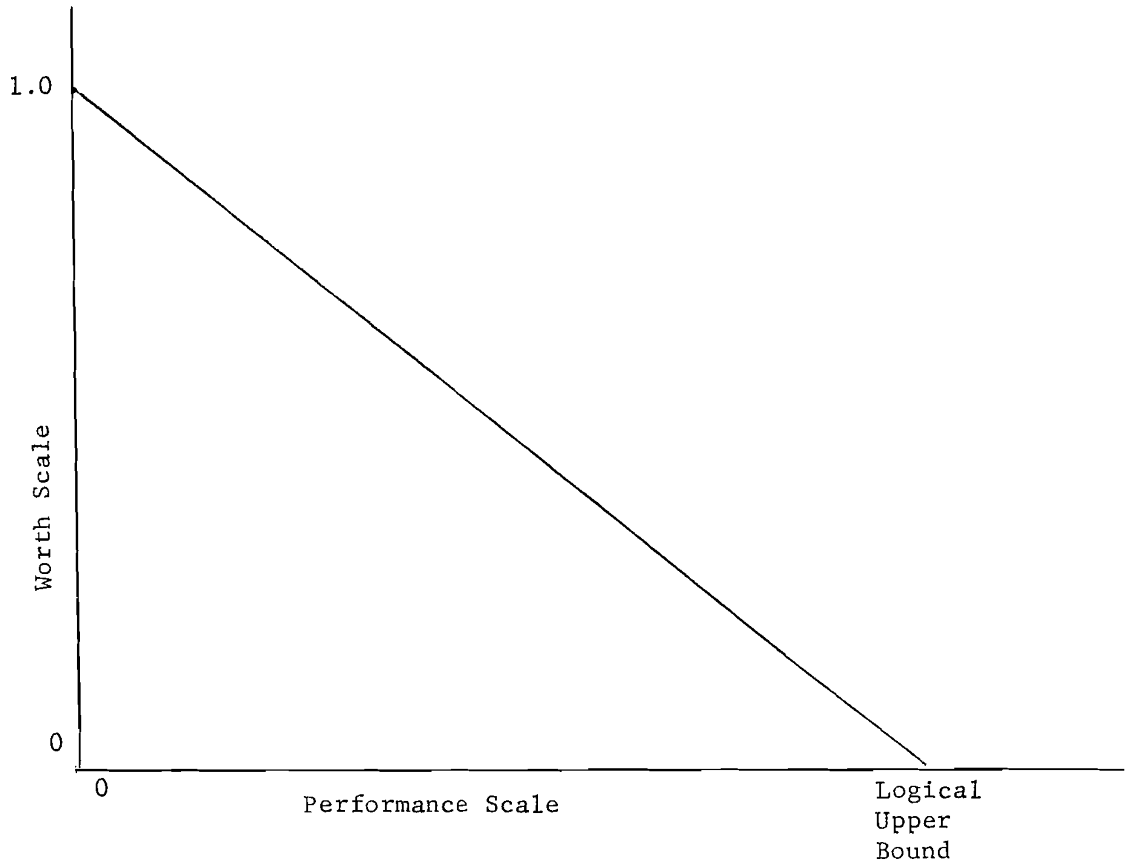


Figure 6

This completes the procedure.

Appendix L

SCORING PROCEDURE 12

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance; and
7. uniformly accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 7.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with negative second derivative (indicating uniform acceleration).

2. The maximum of the quadratic function falls exactly at the point whose coordinates are (performance = zero, worth score = one).

3. The quadratic function falls to the point whose coordinates are (performance = logical upper bound, worth score = zero).

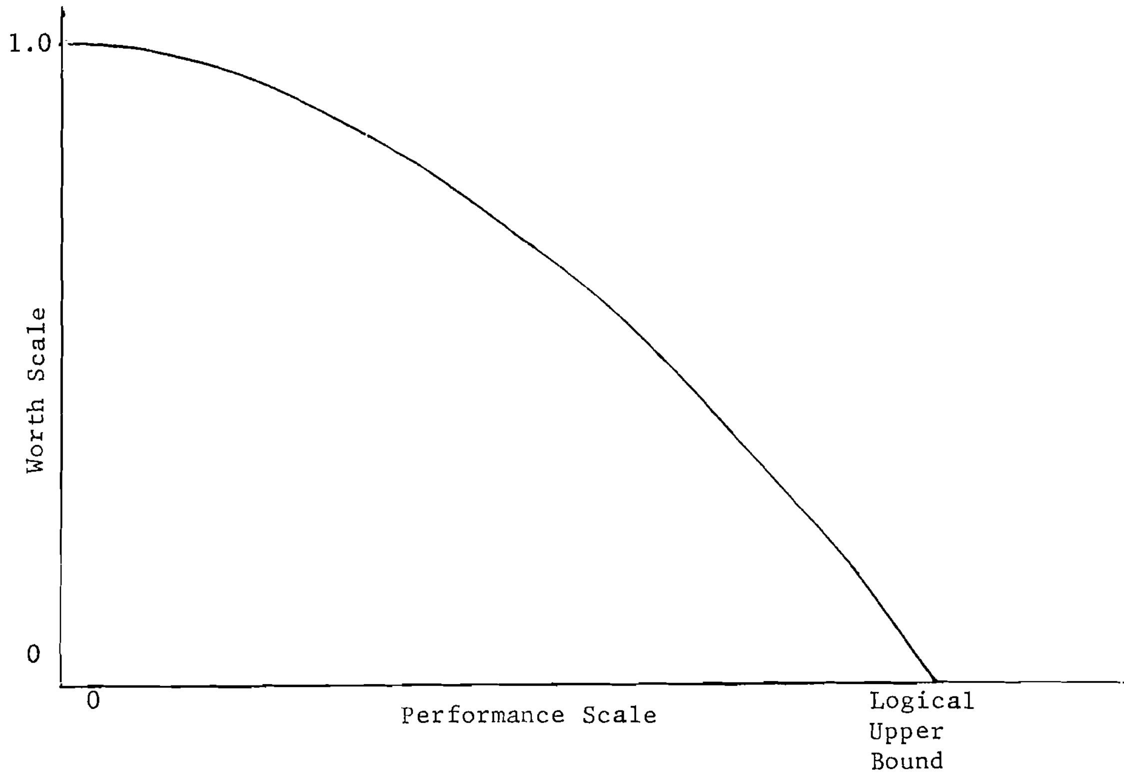


Figure 7

These three assumptions completely determine a scoring function (see Fig. 7) whose equation is

$$\text{worth score} = 1 - \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2.$$

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix M

SCORING PROCEDURE 13

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance; and
7. uniformly accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 8.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with positive second derivative (indicating uniform acceleration).

2. The minimum of the quadratic function falls exactly at the point whose coordinates are (performance = logical upper bound, worth score = zero).

3. The upper left-tail of the function passes through the point whose coordinates are (performance = zero, worth score = one).

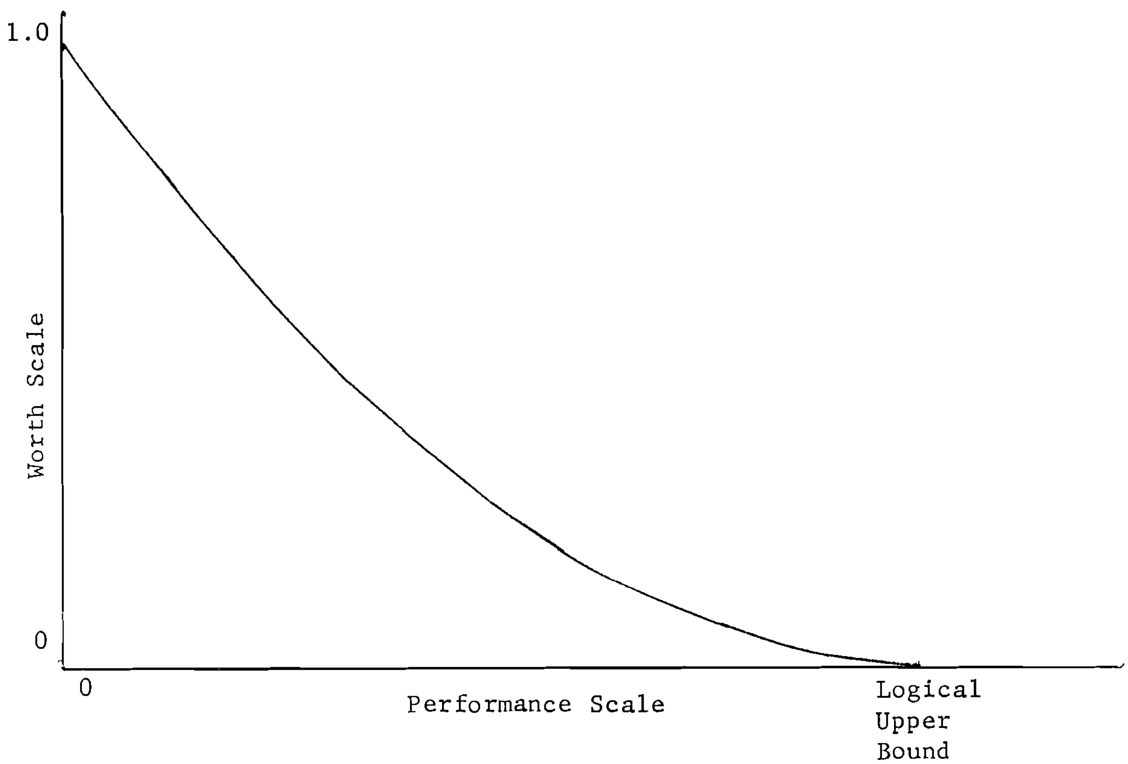


Figure 8

These three assumptions completely determine a scoring function (see Fig. 8) whose equation is

$$\text{worth score} = 1 - 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) + \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2 .$$

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix N

SCORING PROCEDURE 14

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance; and
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 9.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = \tfrac{1}{2} + \tfrac{1}{2}\,\text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right],$$

where $\pi = 3.1416$,
and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

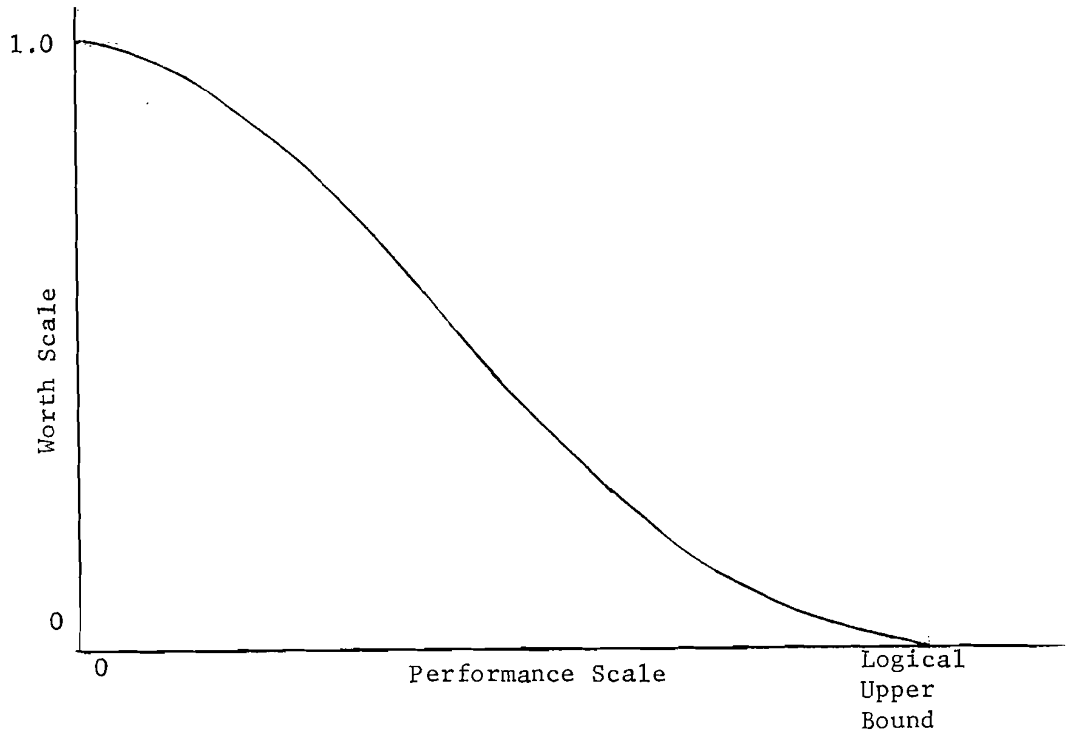


Figure 9

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix O

SCORING PROCEDURE 15

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance; and
7. first decelerating, then accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 10.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = \frac{3}{2} - 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) - \tfrac{1}{2}\,\text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right],$$

where $\pi = 3.1416$,
and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

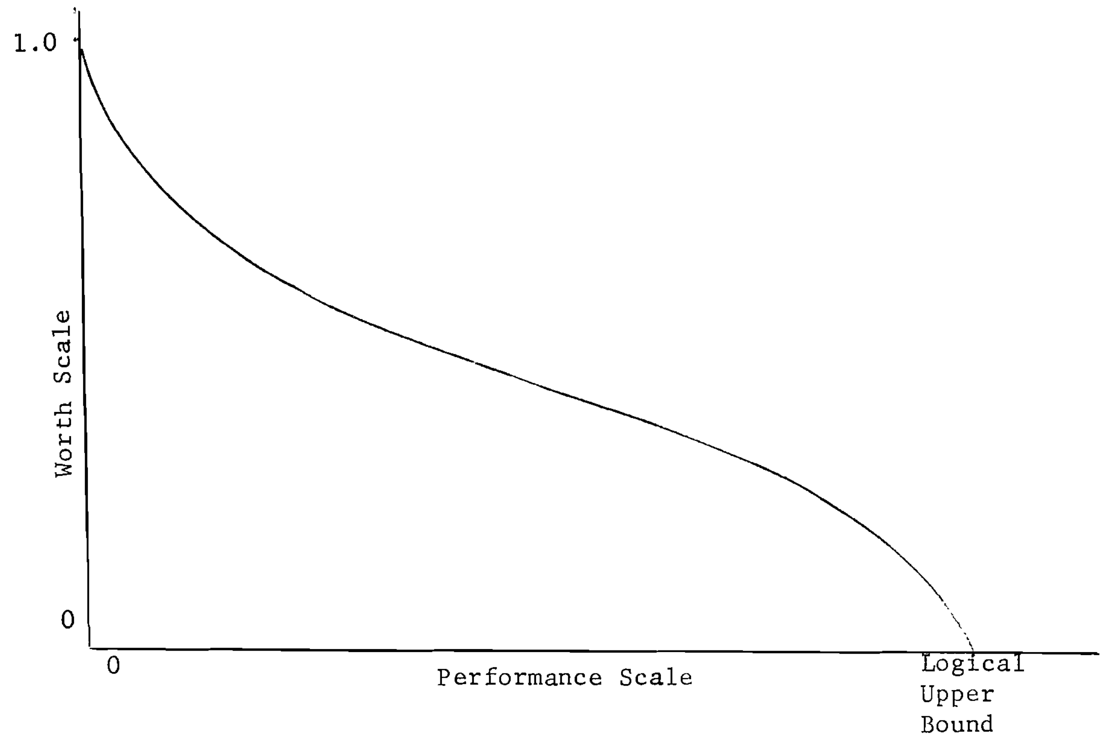


Figure 10

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

Appendix P

SCORING PROCEDURE 16

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to infinite performance; and
7. uniformly decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 11.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = 1 - \exp\left[(-k)\ (\text{measured performance})\right]$$

where exp is the exponential function with basis e = 2.718, and k is a positive fitting constant,

still, the exact value of the fitting constant cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

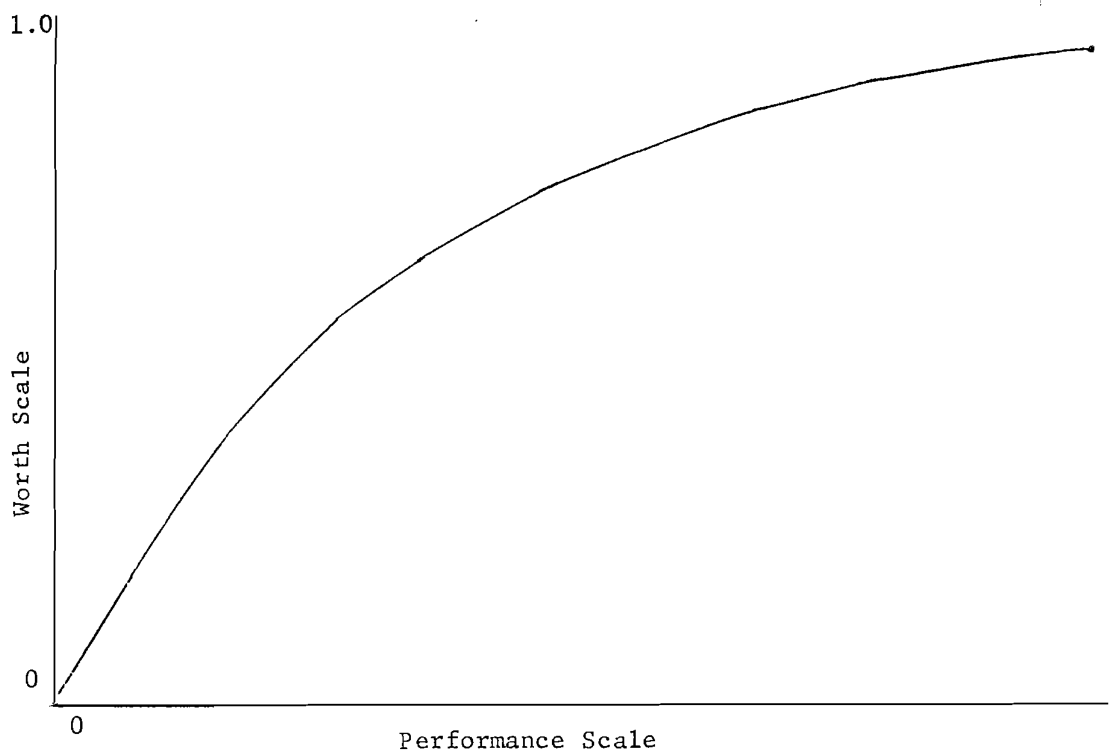


Figure 11

Appendix Q

SCORING PROCEDURE 17

References in text: Section VII, subsection Mapping Out Scoring Functions

The performance measure under scrutiny has been determined to have the following charaacteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to infinite performance; and
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears in Fig. 12.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = \exp\left[(-a)\ \ (\text{measured performance})\right]^{-b},$$

where exp is the exponential function with basis e = 2.718, and both a and b are positive fitting constants ($b \geq 1$),

still, the exact values of the fitting constants cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

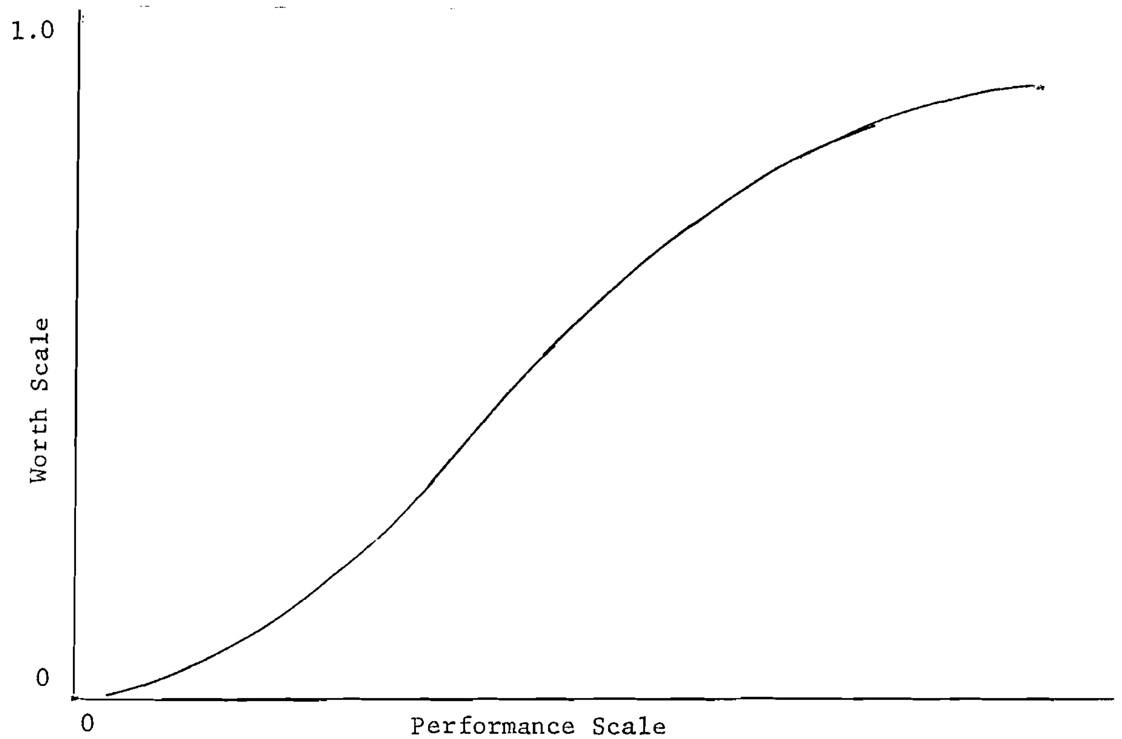


Figure 12

Appendix R

SCORING PROCEDURE 18

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. reverse preference relationship;
5. worth score zero assigned to infinite performance;
6. worth score one assigned to zero performance; and
7. uniformly decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 13.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = \exp\left[(-k)\ \ (\text{measured performance})\right],$$

where exp is the exponential function with basis $e = 2.718$, and k is a positive fitting constant,

still, the exact value of the fitting constant cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

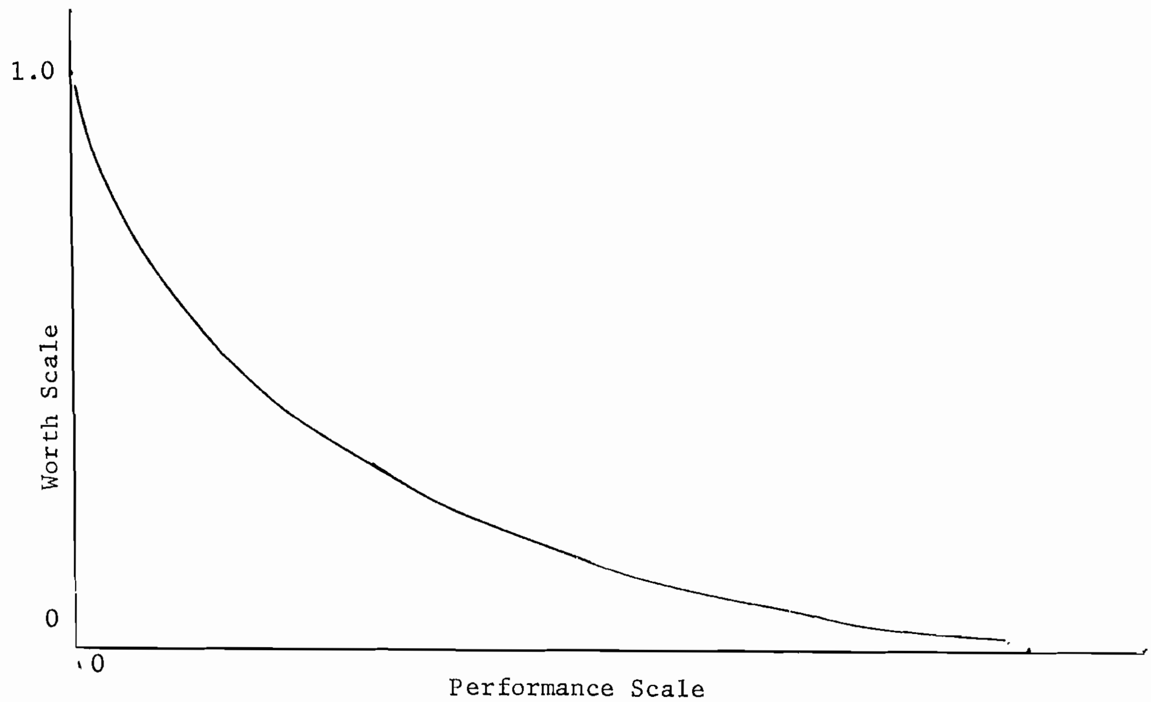


Figure 13

Appendix S

SCORING PROCEDURE 19

References in text: Section VII, subsection Mapping Out Scoring Functions.

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. reverse preference relationship;
5. worth score zero assigned to infinite performance;
6. worth score one assigned to zero performance; and
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Fig. 14.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = 1 - \exp\left[(-a)\quad(\text{measured performance})\right]^{-b},$$

where exp is the exponential function with basis e = 2.718, and both a and b are positive fitting constants ($b \geq 1$),

still, the exact values of the fitting constants cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

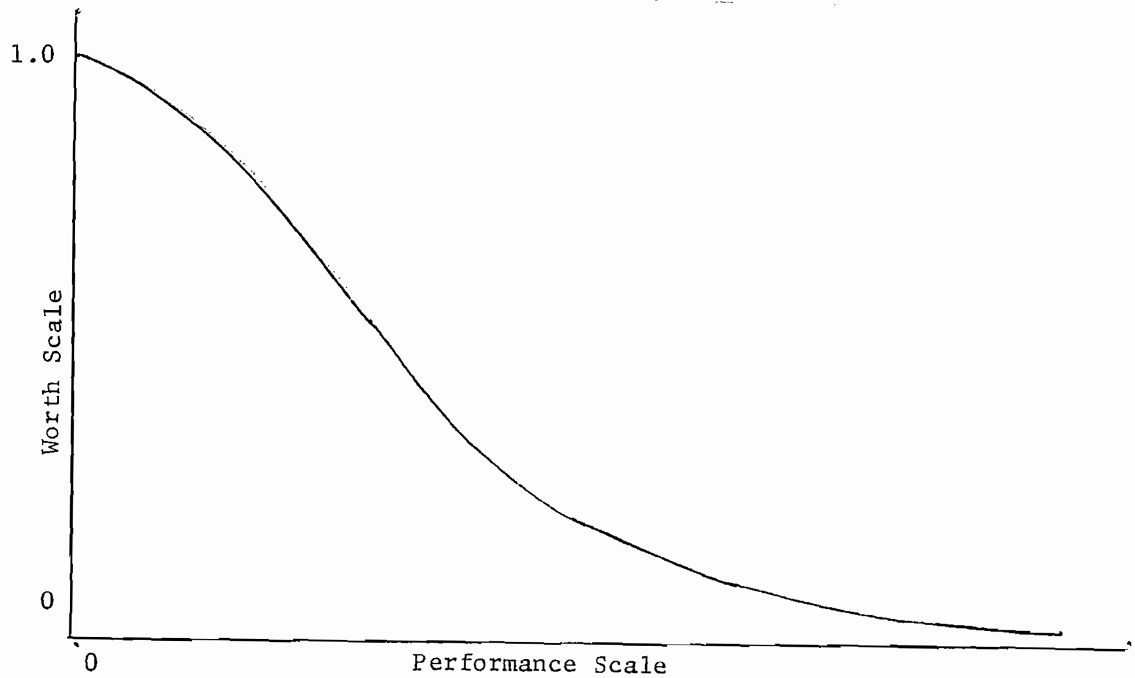


Figure 14

Appendix T

SCORING PROCEDURE 20

Reference in text: Section VII, subsection Mapping Out Scoring Functions.

This procedure constitutes a continuation of each of the previous procedures listed below:

1. Procedure 4
2. Procedure 7
3. Procedure 8
4. Procedure 9
5. Procedure 10
6. Procedure 12
7. Procedure 13
8. Procedure 14
9. Procedure 15
10. Procedure 16
11. Procedure 17
12. Procedure 18
13. Procedure 19

The general shape of the scoring function to be formulated has already been determined and inspected visually in one of these previous procedures. The purpose of this procedure is to select a specific curve of the general shape already determined.

Step 1. Assemble one or more decision makers.

Step 2. Prepare a standard sheet of graph paper for each decision maker laid out and marked off in the following manner.

1. Lay the worth scale along the vertical axis of a Cartesian coordinate plane.

2. Mark off zero worth points at the origin of the graph and one worth point on the vertical axis near the top of the graph.

3. Mark off tenths of a point at equally-spaced intervals along the vertical axis between zero and one worth point.

4. Lay the performance scale along the horizontal axis.

5. Mark off zero performance at the origin of the graph and either the logical upper bound (if one exists) or some amount of performance substantially in excess of (say 50 percent greater than) the anticipated maximum proposed performance on the horizontal axis near the right-hand edge of the graph.

6. Establish convenient, equally-spaced performance subdivisions along the horizontal axis, and mark these off.

Step 3. Each decision maker will then ask himself the following question. "What level of performance, if promised by an alternative, should be considered ten percent successful in satisfying the related lowest level performance criterion?" Indicate this level of performance by placing an "x" in the interior of the graph at the position corresponding to that estimated level of performance along the horizontal performance scale and the ten percent or one-tenth worth point level along the vertical worth scale.

Step 4. Repeat Step 3 for the twenty percent, thirty percent, forty percent, fifty percent, sixty percent, seventy percent, eighty percent, and ninety percent worth point levels, respectively.

Step 5. Each decision maker should now have on his sheet of graph paper nine "x" marks. If the performance measure for which a scoring function is being formulated possesses a logical upper bound, two additional "x" marks may be placed on the graph -- one at zero performance, and the other at the logical upper bound. If the performance measure possesses no logical upper bound, only one additional "x" mark may be placed on the graph corresponding to zero performance. Place the additional "x" mark(s) on the graph.

Step 6. Collect the graphs from each separate decision maker. Compute the (possibly weighted) arithmetic mean (averaged over separate decision makers) for each of the nine percentage levels along the worth scale.

Step 7. Prepare a new sheet of graph paper identical to the sheets prepared in Step 2.

Step 8. Plot the nine average points computed in Step 6 on this new sheet prepared in Step 7.

Step 9. With the aid of a French curve, draw a smooth curve of the predetermined general shape through the average points plotted in Step 8. The result is a scoring function in graphical form.

Step 10. To use this graphical scoring function, note the actual amount of performance promised by an alternative, and read the corresponding point score directly off the graph.

This completes the procedure.

Appendix U

AN EXPERIMENTAL TEST OF THE ASSESSMENT PROCEDURE BY PROFESSIONAL DECISION MAKERS

In the preceding sections of this paper, a systematic procedure to aid in the assessment of worth was first developed and then illustrated. The purpose of this procedure, it will be recalled, is to help decision makers formulate and articulate a consistent assessment structure (really a complex algorithm) for assessing the worth of specified alternatives in a definite choice situation. Once formulated, this assessment algorithm may be applied to each specified alternative so as to generate a numerical index of its overall worth.

The experiment, whose results will be reported in this Appendix, was designed to test the assessment procedure--that is, to determine whether or not the procedure could be implemented by professional decision makers and, if so, with what degree of success.

A BRIEF REVIEW OF THE WORTH CONCEPT

In order to recall the conceptual foundations of the assessment procedure and to motivate discussion of the experiment, five critical assumptions about the worth concept are restated below.*

1. Worth is an internal property of human beings. Worth notions exist within the perceptual and attitudinal apparatus of human decision makers--not as external properties of the physical objects and activities which human beings assess and to which they impute worth. To assess the worth of an object or activity, therefore, is to measure a decision maker's response (e.g., verbal assessment, behavioral choice, etc.) to that object or activity.

2. In general, human notions of worth are multidimensional rather than unidimensional. This means two things:

(a) A given physical object or activity is perceived as relevant simultaneously to more than one human objective.

---

*The reader is referred to Section III of this paper for a more complete discussion of the worth concept.

(b) A given human objective may be satisfied by more than one alternative object or activity.

3. An individual's notions of worth need not necessarily be shared by others (i.e., consensual validation is not a definitional requirement of legitimate worth notions), although some consensus can be expected, particularly within his reference group.

4. An individual's notions of worth need not necessarily be stable over time (i.e., temporal stability is not a definitional requirement of legitimate worth notions), although some stability can be expected, particularly where his more important values are concerned.

5. Worth notions do not usually exist in a conscious, clearly defined, and logically structured form within the minds of human decision makers. However, with some effort, a consistent assessment structure can be formulated to reflect an individual's notions of worth, so long as certain practical limitations on the ability to conceptualize are observed.

A BRIEF REVIEW OF THE ASSESSMENT PROCEDURE

The assessment procedure, it will be recalled, involves several sequential operations.*

1. Assuming that a job to be done and/or a set of activities to be performed has been described, formulate a list of overall job objectives by abstraction from the job description.

2. Refine each higher-level objective in terms of two or more lower-level, independent performance criteria which define more precisely what is intended by or subsumed under the meaning of the higher-level objective. Generate thereby a complete criterion hierarchy.

3. Interpret lowest-level criteria in terms of physical performance measures.

4. Specify individual worth relationships perceived as holding between each lowest-level criterion and its linked performance measure.

*Review Section VII for a more complete presentation.

5. Establish an overall index of worth, considering all of the previously listed objectives and sub-criteria simultaneously.

If a decision maker can successfully complete the above five operations, he will have created an assessment structure (really a complex algorithm) by means of which a single cardinal worth number may be assigned to any specified alternative in a given choice situation. Inputs to this assessment algorithm consist of various physical performance measures selected by the decision maker as describing the relevant measurable attributes of an alternative. The output of this assessment algorithm is a single cardinal number purporting to represent the worth imputed by the decision maker to that alternative.

THE PURPOSE OF THE EXPERIMENT

As stated previously, the purpose of the experiment was to test the assessment procedure. In particular, the following questions were raised concerning the impact of the procedure upon professional decision makers as they develop preferences for specified alternatives and eventually choose one of them.

1. Are professional decision makers both able and willing to undertake the complete assessment procedure in making a choice among specified alternatives?

2. If so, which aspects of the procedure are difficult to interpret and implement?

3. Does introduction of the procedure into the decision making process serve to clarify, to confuse, or to have no noticeable impact upon individual preferences for alternatives? If there is a noticeable impact, how great is it?

4. Does the procedure increase, decrease, or have no noticeable impact upon the number of preference discriminations made by decision makers among alternatives? If there is a noticeable impact, how great is it?

5. Does the procedure increase, decrease, or have no noticeable impact upon a decision maker's confidence in the accuracy of his indicated preferences?

6. How satisfied are decision makers with using the procedure? Specifically, do they consider it helpful in improving the quality of their final choices? If so, by how much?

7. Does implementation of the procedure serve to alter preferences for alternatives? If so, by how much and in what ways?

8. How aware are decision makers of the extent to which the procedure serves to alter their preferences?

9. To what extent do decision makers feel that any gains made in clarification, confidence, satisfaction, and/or appropriate alteration are worth the additional costs in time and effort expended (by implementing the procedure) to realize these gains?

10. To what extent and in what ways does implementation of the procedure serve to alter attitudes on the part of decision makers toward formal, quantitative techniques of assessment?

11. To what extent will decision makers spontaneously adapt various aspects of the procedure to other decision situations lying beyond the scope of the experiment itself (e.g., to situations more closely resembling the real world)?

These questions constitute the specific senses in which validation of the procedure were sought experimentally. The experiment itself, important results, and overall conclusions will be reported subsequently in summary form.*

THE CONTEXT OF THE EXPERIMENT

Several years ago, the Department of Defense established a school at Wright-Patterson Air Force Base to train military and Civil Service personnel in the intricacies of modern weapons systems management. Military officers from all three branches of the Armed Services and Civil Service personnel from various defense-oriented government agencies (e.g., the National Aeronautics and Space Administration)

---

*A more complete exposition of the experiment can be found in James R. Miller III, The Assessment of Worth: A Systematic Procedure and Its Experimental Validation, Doctoral Dissertation, Massachusetts Institute of Technology, September 1966.

are selected four times each year to participate in an eleven-week training course. A class consists of approximately sixty such individuals holding the rank of Colonel, Lieutenant Colonel, Captain (Navy), Commander, GS-14, GS-15, or the equivalent, and with at least some (in most cases, substantial) prior experience managing government projects. Since the purpose of the course is to train project managers, a large part of the curriculum is devoted to new techniques in scientific management -- particularly those espoused by the Department of Defense. The eleventh and final week of the course consists of a computer-simulated game played by teams of five participants each. The computer is programmed to simulate contractor responses to various decisions made by each team as it progresses through the design, selection, installation, and eventual operation of a typical weapons system.

This eleven-week training course constituted the context of the experiment. The sixty military and Civil Service personnel being trained for duty as project managers comprised the sample of experimental subjects.

SPECIFIC DESIGN OBJECTIVES

In designing the experiment, the following specific objectives were set forth.

1. First, it seemed essential to select a sample of experimental subjects who regularly make important decisions among complex alternatives. After all, it is for precisely this kind of person that the assessment procedure was primarily (if not exclusively) designed. It is not clear that other kinds of people would be either willing or able to undertake such an arduous task.

2. Second, it seemed desirable to have each subject make a definite and clearly observable decision (i.e., choice among alternatives) concerning some issue which he would regard as meaningful and whose consequences would be directly and visibly related to his future well-being. By requiring each subject to make an observable choice, experimental measures of preliminary preferences (for the various alternatives) could be formulated and later tested for their ability

to predict his final choice. By selecting an issue which he would perceive as both meaningful and bearing directly upon his future well-being, each subject could be expected to expend a reasonable amount of time and effort in formulating an assessment structure and applying this to the alternatives.

3. Third, to remain compatible with the assessment procedure, the choice had to be constrained to a fixed set of discrete and clearly specified alternatives.

4. Fourth, it seemed desirable to have all sixty subjects assess similar alternatives in an identical decision situation. This would permit making valid comparisons of results across subjects.

5. Fifth, it seemed desirable to have all sixty subjects make their assessments independently of one another. The focus of this experiment was upon individual (as opposed to group) decision making processes.

6. Sixth, it seemed desirable to make the decision situation relatively simple, relatively familiar, and restricted to a manageable number of alternatives. This would serve to economize time and effort both on the part of the experimenter and on the part of the experimental subjects (no prior training required).

7. Finally, to provide bases against which results of the complete assessment procedure could be compared, it seemed desirable to design experimental manipulations in such a way as to obtain similar measures of the impact of three alternative modes of assessment. These were:

A. Spontaneous assessment with neither explicit information about the alternatives nor any explicit guidance on how to assess their worth or how to make a final choice;

B. Assessment with the aid of raw information about the alternatives, but without any systematic guidance on how to utilize such information in assessing their worth or arriving at a final choice; and

C. Partially guided assessment (i.e., the first part of the complete procedure developed in Section VII including only those operations designed to generate a criterion hierarchy

and performance measures, but excluding the subsequent scoring and weighting operations).

THE DECISION SITUATION, THE ALTERNATIVES, AND THE FINAL CHOICE

Recall that all sixty experimental subjects form teams of five participants each at the end of the training course. Through the medium of a computer game against simulated defense contractors they then proceed to test their newly-acquired knowledge. For purposes of the experiment, the decision which each individual subject had to make was to choose partners and thereby form a team to play the computer game.

Assuming five-man teams (of which there were twelve), an alternative consisted of a group of four other participants who, along with the individual making the choice, could constitute a complete five-man team.

If each subject were permitted to choose any four partners from among the entire remainder of the class, then he would have to consider over 455,000 alternative teams. This was obviously too many for any one person to handle. Consequently, a series of experimental devices had to be employed in order to reduce the alternatives to a manageable number.

The first device was to subdivide the sixty subjects into six sub-groups of ten each. Subdivision was performed prior to the beginning of the training course with the aid of a random number table. Then, each subject was asked to peruse a list of ten names (including his own) and to subdivide the remaining list of nine others into two sub-lists. The first sub-list contained six names of preferred candidates for inclusion in his final team, while the second sub-list contained the three remaining names. Subjects were asked to perform this latter subdivision after having had a few days to acquaint themselves with other participants in the training course. By means of these two devices, each subject then had only six other candidates from whom to choose four team partners. This served to reduce the number of alternative teams which each individual must consider to fifteen.

However, despite these experimental devices, there still remained the problem of giving each subject an independent choice to make. Except by unlikely accident, not every individual in a ten-man sub-group could have his complete choice of partners fulfilled. If two or more individuals included the same third individual in their most preferred team, but failed to include each other, then somebody would have to lose. Consequently, a third experimental device had to be employed to obviate this difficulty and to maintain the prospect of an independent decision for all sixty subjects. It was decided to announce at the outset of the experiment that one subject in each of the ten-man sub-groups would have his choice of team partners honored. The remaining five subjects not chosen by him would be grouped to form a second team. Exactly whose choices were to be honored remained undetermined until the end of the experiment, and a random number table was used to make this determination at that time. Therefore, each subject might proceed on the assumption that he would be making the final choice, for his chances would be just as good as anyone else's of having his choice honored.

THE EXPERIMENTAL PROCEDURE

Prior to the beginning of the experiment, all sixty subjects were assigned at random to three groups. Each group contained two of the ten-man sub-groups, making twenty subjects in all. One of these twenty-man groups performed the complete assessment procedure developed in Section VII. The second group performed part of the procedure (up to the point of generating a criterion hierarchy and selecting performance measures). The third group received only raw information about their alternatives, but no systematic guidance concerning its utilization. All three groups performed initial operations designed to measure the impact of neither information nor guidance.

A battery of written questionnaires, in conjunction with a schedule of personal interviews, was designed and administered during the first ten weeks of the training course. Through these

instruments data were gathered concerning the impact of the various assessment procedures upon the decision making process. At the end of the tenth week, each subject made his final choice of team partners. Five-man teams were then formed on the basis of these choices, and all sixty subjects participated in the computer simulation exercise during the eleventh and final week of the training course.

SATISFACTION OF THE SPECIFIC DESIGN OBJECTIVES

The first objective -- testing out the assessment procedure on professional decision makers -- was satisfied by the particular choice of experimental subjects and the experimental context. All sixty participants in the training course are sent to the school for the express purpose of receiving education in decision making. Most of them have had extensive practical experience in assessing and choosing among complex alternatives prior to coming. The curriculum focuses heavily upon decision making techniques, and the work-pace is intensive. Students live on the Air Base throughout the eleven-week period and are required to attend six hours of class each day. Consequently, on the basis of these personal background and contextual factors, it seemed reasonable to hope that both the subjects and the setting would provide an appropriate vehicle for validating the assessment procedure.

The second objective -- having each subject make a definite and clearly observable decision -- was satisfied by requiring everyone to choose four team partners at the end of the experiment, just prior to playing the computer game. The choice was definite. It was clearly observable by the experimenter (although not by the subject's fellow students). It could have an immediate impact upon his chosen team's performance in the game itself. Since the game was advertised in advance as competitive, and since previous participants in the game had demonstrated substantial personal commitment and competitive zeal, it was reasonable to hope that subjects would take the experimental decision seriously.

The third objective -- providing a fixed set of discrete alternatives -- was satisfied by means of the first two experimental devices.

Each subject had a fixed set of fifteen discrete teams from which to make a final choice (i.e., the fifteen logically possible combinations of six team partners taken four at a time).

The fourth objective -- having all sixty subjects assess similar alternatives in an identical decision situation -- was likewise satisfied by these two experimental devices.

The fifth objective -- inducing each subject to make an independent decision -- was satisfied by the third experimental device. By announcing in advance that an individual's choice of team partners would be honored if and only if his name were selected by a completely random mechanism and without regard to whom he chose or who chose him, it was hoped to discourage the formation of coalitions and the adoption of competitive bidding strategies. In addition, it was decided to give continual instructions to the subjects requesting that they refrain from discussing with one another their preferences, their assessment criteria, or their anticipated final choices.

The sixth objective -- presenting a relatively simple and familiar decision situation -- was satisfied by the nature of the required choice. Choosing partners for some group enterprise is a familiar decision made many times in almost everyone's lifetime. Choosing up sides for an athletic contest or parlor game, selecting new members for a social or business organization, and choosing a marriage partner are common examples.

Satisfaction of the seventh design objective -- providing bases of comparison -- was achieved by splitting the class into three groups of twenty each and having them undertake different modes of assessment.

SUMMARY OF RESULTS AND CONCLUSIONS

The experiment yielded the following results and conclusions.

1. The complete assessment procedure developed in Section VII was implemented in its entirety by all twenty of the subjects introduced to it. However, one subject stated in advance that he viewed the procedure as an empty ritual. His implementation was, therefore, only nominal and signified no real commitment to its overall intent.

A second subject chose to substitute alternative procedures of his own design for the scoring and weighting operations suggested in Section VII. From these results it was concluded that the procedure could be implemented by professional decision makers.

2. The one aspect of the procedure which consistently produced confusion and misunderstanding was the issue of independence among objectives and performance criteria. It required a fair amount of interpretive discussion to clarify the meaning of this concept. Hence, it was concluded that further efforts might profitably be expended upon this portion of the procedure.

3. The complete assessment procedure was judged superior to all three of the alternative modes of assessment included in the experiment. In addition, subjects introduced to the complete procedure tended to adapt it to another decision context (i.e., to making decisions during the course of the computer simulation exercise) to a significantly greater extent than did subjects introduced to the alternative modes of assessment.

4. Almost every conscious assessment activity which subjects perceived as relevant to making their choices served to clarify their preferences for alternatives. In particular, receiving factual information, being required to articulate and structure assessment criteria, and being required to quantify their preferences all had this effect. The mere realization that a choice had to be made, accompanied by preliminary efforts to structure the alternatives, had the same effect. However, when subjects did not perceive such activities as relevant, even though they were alleged to be, clarification did not occur. When clarification did occur, its magnitude varied with the particular type of activity engaged in. Of critical importance were those kinds of activity which challenged and thereby tested the validity of existing preferences (e.g., comparison of informal, subjective preferences with formally derived quantitative outputs of the assessment algorithm).

5. The number of preference discriminations spontaneously made by subjects among alternatives depended primarily upon individual factors. Changes thereto induced under alternative modes of assessment also depended upon individual factors.

6. Almost every conscious assessment activity perceived as relevant to the decision served to increase confidence in the accuracy of stated preferences. In particular, the four modes of assessment designed into the experiment had this effect. Irrelevant activities did not have this effect. Again, the magnitude of this effect depended upon the particular type of activity.

7. The same results concerning clarification occurred in the case of satisfaction derived by subjects from undertaking various modes of assessment. Satisfaction, here, refers to the degree to which such activities were perceived as helpful to improving the quality of the final choice.

8. Although subjects did receive clarification, satisfaction, and additional confidence from undertaking various modes of assessment, this did not guarantee that they would overtly alter prior preference commitments in light of newly-perceived implications. Once again, provision of a challenge or validity check (e.g., comparison of subjective preferences with numerical outputs of the assessment algorithm) was of critical importance. When such checks were performed, then overt commitment generally did follow.

9. On the other hand, changes in preference occurred covertly following almost every conscious and relevant assessment activity, but did not occur (apart from random instability) unless such activity was perceived as relevant. The magnitude of such changes decreased steadily as confidence and clarification increased and as the moment of final decision drew near.

10. Without knowing precisely what their previous preferences were, subjects tended to underestimate their temporal stability. They also tended to underestimate the magnitude of changes in stability over time. Both of these phenomena became less pronounced if they made a definite and overt commitment to a particular preference structure.

11. The perceived value of engaging in various assessment activities compared to the time and effort expended depended critically upon the type of activity engaged in. When the activity was perceived as irrelevant, it was considered a waste of time. However, even when

the activity was perceived as relevant, it was not always considered sufficiently valuable to justify the time and effort expended. Once again, providing a challenge or validity check was particularly important in this respect.

12. The complete assessment procedure developed in Section VII had a four-fold overall impact upon decision processes.

(a) Its primary impact was to induce subjects to formulate and validate a consistent assessment structure. Validation was provided by comparing formally derived with subjective preferences, and the quantitative aspects of the procedure were critical in this respect. Alternative modes of assessment, lacking quantitative aspects, did not, in general, produce this effect.

(b) In the process of formulating an assessment structure, preferences for alternatives were significantly altered. However, they were not altered randomly, but rather in a manner directed toward the final choice.

(c) When the entire procedure was followed, particularly the final steps of quantitative assessment, a mechanism was provided to validate preferences. This, in turn, induced favorable reactions to formal assessment techniques. It also induced at least intermediate-term changes both in attitudes toward the procedure and in preferences for alternatives. On the other hand, when only part of the procedure or none of the procedure was followed, the reaction of subjects was nowhere near as favorable nor as permanent.

(d) Another important impact was to measure and display assessment criteria, which can be useful both for purposes of normative decision making and for purposes of scientific description. The alternative modes of assessment did not produce this result -- at least not to the same extent.

BIBLIOGRAPHY

Miller, J.R. III, "The Assessment of Worth: A Systematic Procedure and its Experimental Validation," Ph.D. Thesis, Massachusetts Institute of Technology, 1966.

Pardee, F. S., et al., Measurement and Evaluation of Transportation System Effectiveness, The RAND Corporation, RM-5869-DOT, 1969 (forthcoming).

Raiffa, H., Preferences for Multi-Attributed Alternatives, The RAND Corporation, RM-5868-DOT/RC, 1969.

von Neumann, J., and O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, New Jersey, 1947.